Charles University

Faculty of Science

Ph.D. Study Program: Botany



Mgr. Mária Šurinová

Application of molecular methods in population genetic studies

Aplikace molekulárních metod ve studiích populační genetiky

Doctoral thesis/Dizertační práce

Supervisor/Vedoucí práce: Prof. Zuzana Münzbergová Ph.D.

Praha, 2021

Prohlášení autora / Author's declaration:

Prehlasujem, že som záverečnú prácu spracovala samostatne a že som uviedla všetky použité informačné zdroje a literatúru. Túto prácu ani jej podstatnú časť som nepredložila k získaniu iného nebo rovnakého akademického titulu.

I hereby declare that I made this thesis independently, using only the mentioned references. I did not submit this thesis nor its part for any other degree or diploma.

V Praze dne 16. 11. 2021 Mária Šurinová

Acknowledgements

First, I would like to thank my supervisor, Zuzana Münzbergová, for everything she has taught me, for her help and patience especially during writing process, for stimulating discussions with her and for all the opportunities she gave me. Next, I would like to all co-authors of my papers and institutional colleagues: Andrea Jarošová, Zuzana Líblová, Viera Hadincová, Tomáš Dostálek, Hana Pánková, Jana Rydlová, David Püschel, Eva Níčová, Dana Parysová and Ilona Jarošincová. Also I would like to thank Štepán Stočes and Petr Vácha from Seqme Company for their help far beyond their responsibilities.

Last, I would like to thank my family, especially to Jirko and my friends Radka Pavlíčková-Brabcová, Tereza Balšánková and Veronika Olivová for constant support and enthusiasm.

Contents

1. Abstract **6**

Abstrakt **8**

2. Introduction **10**

2.1. Polyploidization 10

2.1.1. Polyploidization and adaptation 12

2.2. Fragmentation of populations 14

2.2.1. Species response to fragmentation 15

2.3. Overview of my research 19

2.4. Author contribution statement 23

2.5. References 25

Study 1 ..... **34**

Study 2 ..... **47**

Study 3 ..... **61**

Study 4 ..... **70**

Study 5 ..... **83**

Study 6 ..... **88**

Study 7 ..... **99**

Study 8 ..... **112**

3. Conclusions **115**

1. Abstract

A vast range of factors shape the genetic structure of plant populations. In this thesis, I focus on two of them. The first factor, polyploidization, is a process of chromosome set multiplication through whole-genome duplication within a single species (autopolyploids) or hybridization of two different species (allopolyploids). It rapidly brings changes into genomes, allowing species to occupy distinct niches, adapt to new habitats, colonize them, or adapt to changing environment in their native range. But it comes at a certain cost - difficulties in mitosis and meiosis, changes in cellular architecture. Furthermore, after cytotype establishment, new individuals have to deal with cytotype exclusion effect, competition with parental individuals and higher nutrition requirements. In this thesis, I present the effect of polyploidization on populations of three species- *Arabidopsis arenosa*, *Aster amellus* and *Festuca rubra*.

The second factor changing the population genetic structure presented in this thesis is the fragmentation of populations. Population fragmentation can be caused by natural or anthropogenic activities and often leads to overall reduction in population size and reduced connectivity among fragments. Restricted gene flow may threaten long-term population survival due to inbreeding depression and loss of genetic diversity. The genetic structure of fragmented populations can be affected by factors like the number, size and spatial distribution of populations and time since fragmentation. In this thesis, I present the effect of fragmentation on populations of three species- *Gentianella praecox* subsp. *bohemica*, *Minuartia smejkalii* and *Rumex alpinus*.

This thesis aimed to 1, compare meiotic genes between diploid and tetraploid cytotypes of *Arabidopsis arenosa* 2, elucidate the relationship between diploid and hexaploid cytotypes of *Aster amellus* 3, describe effects of temperature and moisture on genetic diversity of hexaploid *Festuca rubra* populations 4, characterize the genetic structure of *Gentianella praecox* subsp. *bohemica* remaining populations 5, investigate population genetic structure of *Minuartia smejkalii* survived populations 6, describe the genetic structure of *Rumex alpinus* native populations.

We found 1, selection on meiotic genes and differences in meiotic stabilization between *Arabidopsis arenosa* cytotypes 2, strongly restricted gene flow between *Aster amellus* cytotypes even at the contact zone. We proposed that these cytotypes may be considered two different lineages 3, both abiotic factors represent likely determinants of genetic relatedness in the system and the level of clonal growth of the species *Festuca rubra* 4, significant reductions

in population size do not result in loss of genetic diversity and reduction of fitness within the *Gentianella praecox* subsp. *bohemica* populations 5, high genetic diversity within remaining seven *Minuartia smejkalii* populations, low to moderate among-populations differentiation. 2–3% of the markers show differentiation patterns consistent with divergent selection, suggesting that local adaptation might have occurred. 6, strong clonal propagation in *Rumex alpinus* populations as preferred reproduction strategy of this invasive species.

In summary, genetic diversity influenced by whole-genome duplication allowed all three investigated species to either expand their native range or prosper and reproduce in harsh environment. For the effect of population fragmentation is possible to conclude that population fragmentation surprisingly did not lead to strong diversity reduction in two investigated endangered species. On the other hand, in invasive species *Rumex alpinus* clonal reproduction strategy led to the low genetic variability of isolated populations.

Abstrakt

Velké spektrum faktorů může formovat genetickou strukturu populací rostlin. V této disertační práci se soustředím na dva z nich. První faktor, polyploidizace, je proces množení chromozomových sad prostřednictvím duplikace celého genomu v rámci jednoho druhu (autopolyploidia) nebo hybridizace genomů dvou různých druhů (alopolyploidia). Tento proces rychle přináší nové varianty do genomů, což umožňuje druhům obsadit odlišné arealy, přizpůsobit se novým biotopům a kolonizovat je nebo se přizpůsobit měnícímu se prostředí v jejich původním areálu. Výhody polyploidizace jsou spojeny s nevýhodami v podobě komplikací v mitóze a meióze a změnami v buněčné architektuře. Kromě toho, po vytvoření cytotypu se noví jedinci musí vypořádat s efektem vyloučení cytotypu, konkurencí s rodičovskými jedinci a vyššími požadavky na živiny. V této práci představuji vliv polyploidizace na populace tří druhů – *Arabidopsis arenosa*, *Aster amellus* a *Festuca rubra*.

Druhým faktorem měnícím genetickou strukturu populací prezentovaných v této práci je fragmentace populací. Fragmentace populace může být způsobena přírodními nebo antropogenními aktivitami a často vede k celkovému snížení velikosti populace a snížení konektivity mezi fragmenty. Omezený tok genů může ohrozit dlouhodobé přežití populace v důsledku inbrední deprese a ztráty genetické rozmanitosti. Genetickou strukturu fragmentovaných populací mohou ovlivnit faktory, jako je počet, velikost a prostorové rozložení populací a čas od fragmentace. V této práci představuji vliv fragmentace na populace tří druhů - *Gentianella praecox subsp. bohemica*, *Minuartia smejkalii* a *Rumex alpinus*.

Cílem této disertační práce bylo 1, porovnat meiotické geny mezi diploidními a tetraploidními cytotypy *Arabidopsis arenosa* 2, objasnit vztah mezi diploidními a hexaploidními cytotypy *Aster amellus* 3, popsat vliv teploty a vlhkosti na genetickou diverzitu populací hexaploidní *Festuca rubra* 4, charakterizovat genetickou strukturu populací druhu *Gentianella praecox subsp. bohemica* 5, zkoumat genetickou strukturu populací druhu *Minuartia smejkalii* 6, charakterizovat genetickou strukturu původních populací druhu *Rumex alpinus*.

Hlavní závěry disertační práce jsou 1, identifikovaná selekce na meiotických genech a rozdíly v meiotické stabilizaci mezi cytotypy druhu *Arabidopsis arenosa* 2, silně omezený tok genů mezi cytotypy druhu *Aster amellus* a to i v kontaktní zóně obou cytotypů, 3, oba abiotické faktory ovlivňují genetickou příbuznost populací a úroveň klonálního růstu druhu *Festuca rubra* 4, významné snížení velikosti populace nemá za následek ztrátu genetické diverzity a snížení fitness u populací *Gentianella praecox subsp. bohemica* 5, detekovaná vysoká

genetická diverzita v rámci zbývajících sedmi populací *Minuartia smejkalii*, nízká až střední diferenciace mezi populacemi. 2–3 % diferencovaných markerů poukazují na divergentní selekci, což naznačuje, že mohlo dojít k lokální adaptaci u studovaných populací 6, silná klonální propagace v populacích *Rumex alpinus* byla identifikována jako preferovaná reprodukční strategie tohoto invazního druhu.

Souhrnem, genetická diverzita ovlivněná duplikací celého genomu umožnila všem třem zkoumaným druhům buď rozšířit svůj původní rozsah, nebo prosperovat a rozmnožovat se v nehostinném prostředí. U studií fragmentovaných populací lze shrnout, že fragmentace populace překvapivě nevedla k výraznému snížení diverzity u dvou zkoumaných ohrožených druhů. Na druhé straně u invazního druhů *Rumex alpinus* vedla klonální reprodukční strategie k nízké genetické variabilitě izolovaných populací.

2. Introduction

Plants occur in populations, defined as a collection of individuals belonging to the same species living in the same area. This definition has two aspects, a genetic one (individuals of the same species) and a spatial one (occurrence in the same area). However, populations are neither genetically nor spatially homogenous, and several types of structures can be studied. Genetic structure describes genotype distribution, spatial structure describes changes in density within the population, and age structure describes the relative composition of young and old individuals. These structures influence each other, and also they change with time (Silvertown and Charlesworth, 2001). This thesis mainly focuses on the population genetic structure of different plant species and factors that influence genetic structure within and between populations.

The main factors influencing the genetic diversity of populations are a, genetic drift and related effective population size, b, bottleneck and founder effect c, selection d, mode of reproduction and polyploidy e, inbreeding (Freeland, Kirk and Petersen, 2011). All these processes act on the population level.

However, populations are not isolated. They are connected through migration and gene flow. Migrated individuals can profit from new populations/habitats by avoiding herbivores, the possibility of inbreeding and to gain more free space and resources. However, they can risk difficulties in reproduction and a lack of resources in an unsuitable habitat. Migrated individuals can contribute to gene flow. Gene flow means that an individual can reproduce and contribute to a set of genes in a population. The main factors influencing gene flow are a, habitat barriers and fragmentation, b, metapopulations, c, hybridization d, the ability of the species to migrate (Freeland, Kirk and Petersen, 2011). All these factors shape the genetic structure of populations in different combinations, which are species-, geography- and environmental- dependent. This thesis is mainly focused on two factors connected with population structure, polyploidy and fragmentation of populations, which were investigated on different plant species with different molecular methods.

2.1 Polyploidization

Polyploids are traditionally defined as either autopolyploids, which arise within a single taxonomic species, or allopolyploids, which are the product of interspecific hybridization (e.g. Ramsey and Schemske, 2002). In both cases, the increase in the number of chromosome sets

through whole-genome duplication cause complex biological problems at cellular level such as genomic instability, mitotic and meiotic abnormalities and altered gene expression (Brownfield and Köhler, 2011; Chen and Ni, 2006; Hollister, 2015; Otto and Michalakis, 1998; Page and Hawley, 2003). Specifically, whole-genome duplication significantly affects chromosome segregation in meiosis. In some polyploids, meiotic stability evolves through an increase in pairing partner preferences, whereas in species that retain random segregation of all homologs, lower genome-wide crossover rates can reduce the formation of multivalent associations among multiple homologs in meiosis. To elucidate the meiotic challenges of polyploidization, we compared meiotic genes in diploid and tetraploid populations of *Arabidopsis arenosa* (Paper 1). We sequenced 12 populations, 8–26 individuals from each population and compared meiosis genes evolution on diploid and tetraploid populations. This comparison showed the evidence for selection on structural proteins that mediate sister chromatid cohesion, the formation of meiotic chromosome axes, and synapsis selection act on meiosis genes. Selection act in the polyploid populations but also independently on a smaller subset of meiosis genes in diploids. The tetraploid lineage has sustained selection on more genes, has more amino acid changes in each, and these more often affect conserved or potentially functional sites. These differences suggest, there are more ways in which meiotic stabilization can be achieved by natural selection. In Seear *et al.*, 2020, previously described meiotic genes were sequenced in additional populations (10 autotetraploid *Arabidopsis lyrata* and *A. lyrata*/*A. arenosa* hybrid populations), and 58 alleles were identified. This indicate dynamic population variability in these genes under extreme selection. Moreover, allelic chimerism between alleles originating from *A. lyrata*/*A. arenosa* and diploid/tetraploids were detected, which reveal gene mechanisms for adaptation.

Another challenge establishing newly formed polyploid lie in the competition with their diploid progenitors. This competition is reflected in the minority cytotype exclusion effect (Levin, 1975) and cytotypes competition (Rodriguez, 1996; Yamauchi et al., 2004). These factors lead to dysfunction in newly formed polyploids and contribute to eliminating the new polyploids from the population. Despite all the challenges mentioned above, the coexistence of intra-populational cytotype mixtures are reported frequently (Certner et al., 2017; Duchoslav et al., 2020; Trávníček et al., 2011; Weiss et al., 2002; Zozomová-Lihová et al., 2015). Ecological requirements of polyploids often differ from those of the parental cytotypes, and ecological differentiation of cytotypes in many species was intensely studied (Arrigo et al., 2016; Blaine Marchant et al., 2016; Kiedrzyński et al., 2021; Laport et al., 2016). In contrast, no significant

ecological differentiation was found for *Centaurea jacea* (Hardy and Vekemans, 2001) or *Aster amellus* (Mandáková and Münzbergová, 2006). To elucidate the relationship between cytotypes of *Aster amellus,* which are morphologically indistinguishable, we used phylogenetic analysis (Paper 2) based on sequencing of chloroplast regions *atp*I–*atp*H, *rps*16 3′ exon–*trn*K (UUU) 5′ exon, *rpl*32–*trn*L and *Pet*N–*psb*M, which exhibited high levels of genetic variation, and PsbE–petL with moderate variation and ITS marker (primer pair ITS4- ITS5) with high levels of genetic variation. Phylogenetic analysis (based on Bayesian inference) of chloroplast regions showed mixed connections between the cytotypes, which may correspond to shared ancestral relationships. On the other hand, sequencing and phylogeny of the ITS nuclear region showed segregation between diploid and hexaploid cytotypes, where each cytotype represents a monophyletic lineage probably homogenized by concerted evolution. Polyploidization of diploid *Aster amellus* progenitors led to a change of ecological niche and allowed species expansion. We also characterized population structure with a higher evolution rate-microsatellite markers. In Paper 3, we used seven microsatellite regions at three different spatial scales: (1) in the single known mixed-ploidy population, (2) in populations at the contact zone and (3) in a wider range of populations across Europe. Separation of the cytotypes at all three spatial scales was identified. Moreover, few hexaploid individuals originated in mixed ploidy populations felt into the diploid cluster despite limited gene flow. The recurrent formation of polyploidization can be an explanatory scenario of this finding. Similarly, Eriksson *et al.*, 2017 studied the diploid- tetraploid complex of species *Medicago prostrata*. Tetraploid *M. prostrata* individuals form a new (cryptic) species, characterized by a partially non-overlapping distribution and distinctive alleles (from the M. sativa complex). Partial reproductive isolation based on the absence of gene flow from tetraploid to diploid *M. prostrata* is apparent. Authors suggest that speciation via autopolyploidy have been reinforced by introgression.

2.1.1 Polyploidization and adaptation

Polyploidization can also enhance the adaptive potential of species. Biotic factors as symbiotic and parasitic relationships influence polyploidization in plants. Fungi, bacteria, and roundworms can also have an impact on polyploidy patterns (Goga et al., 2018). Abiotic factors, such as temperature and water accessibility, play a key role in shaping distributions of species and their populations. Because polyploidization generates unique variation, it may give rise to an immediate shift in phenotype and physiological tolerances and result in shifts in ecological tolerances (Husband And and Schemske, 2000; Manzaneda et al., 2015; Ramsey, 2011).

One of the effects of polyploidization is the increase in cell size, which alters the surface to volume ratio (Melaragno et al., 1993) and have physiological consequences such as increased stomata size and lower stomata density (W. Li et al., 1996; Maherali et al., 2009), thicker epidermis (W. Li et al., 1996) and cell wall (Nassar et al., 2008). These changes can affect transpiration, water transport, hydraulic conductivity, photosynthetic rate, antioxidant response and often leads to greater drought tolerance in polyploids than their diploid relatives (e.g. Maherali, Walden and Husband, 2009; Manzaneda *et al.*, 2012, 2015). Recently, the diploid–tetraploid complex of perennial grass *Themeda triandra* was studied, and it was shown that polyploidy appears dominant in hot and dry regions (Ahrens et al., 2020). Under a combination of drought and heat pressure, polyploidy in this species was shown to have a direct effect on seed, dormancy, and seedling characteristics. Tetraploids produced heavier, more viable seed sets with prolonged dormantion, which allowed plants to avoid germination under unfavorable conditions. Furthermore, once they did germinate, tetraploid seedlings were larger and grew faster compared to their diploid relatives (Godfree et al., 2017; Stevens et al., 2020). In contrast, analysis of leaf anatomy in *Chamerion angustifolium* closely packed mesophyll cells and small conduits in the midvein for tolerating water deficiency in diploids. Thus, diploid *C.angustifolium* is more tolerant to drought than hexaploid plants (X. Li and Guo, 2020). The context for these findings brought mapped global distribution of polyploid plants (Rice et al., 2019). Authors claimed that the effect of temperature is mostly indirect, mediated predominantly by variation in plant lifeforms and, to a lesser extent, by taxonomical composition and species richness.

However, in mentioned and many other studies (Brennan et al., 2009; Brito et al., 2016; Karunarathne et al., 2018; Zhang et al., 2019) temperature and water accessibility (and other abiotic factors) are often correlated to each other and to the geographic position of the populations and separation of the different factors and identification of their effects is thus difficult. A unique climatic grid in Norway allows us to distinguish the effect of temperature and precipitations on studied species. In Paper 4, we studied the genetic structure of hexaploid species Festuca rubra originated in this grid and described the effect of temperature and precipitation separately. Four microsatellites regions were used to study in In 443 individuals, and we identified 63 alleles. Also, ploidy levels were tested by using flow cytometry, and hexaploid cytotype was confirmed. In the population from the most extreme wet cold environment, genetic diversity was the lowest, while the number of alleles per individual was the highest. This is an indication that resilient heterozygous individuals can survive in this

locality. Another identified mechanism to deal with cold environment was clonal propagation. Increased clonality towards the coldest and the coldest wettest climates was found. Because of clonality, the possibility of alpine populations adapting to novel conditions is lower compared to the rest of the populations, in which most of the genetic variation was distributed within populations.

2.2 Fragmentation of populations

Fragmentation leads to the isolation of population fractions, which increase the chance of eventual extinction (Rybicki and Hanski, 2013). Populations can experience natural or/and anthropogenic fragmentation. Both can have similar spatial magnitude (rock pools in the desert, river catchments and estuaries vs urban green areas, natural areas disrupted by highways and larger scale deforestation). Variety of spatial characteristics e.g. fragment size (Collins et al., 2009), shape (Groom and Schumaker, 1993), connectivity (Sullivan et al., 2019), heterogeneity (Kotliar and Wiens, 1990) and stability (Hulshof and Spasojevic, 2020) affect fragmented populations. With ecological effects like edge effect (e.g. forest edges often experience increased light, temperature, wind, Wallace *et al.*, 1997), Allee effect (an increase of fitness reduction as population size declines, Stephens, Sutherland and Freckleton, 1999) and kin competition (Hamilton and May, 1977) fragmented populations also have to cope. Smaller populations are more prone to environmental fluctuations (e.g. pathogen outbreaks, fire) and demographic stochastic events (birth and death rate, sex ratios) (Ewers and Didham, 2006; Laurance et al., 2002; Shaffer, 1981). Probably the most important characteristic which affects the dynamic and survival of fragmented populations is the degree to which dispersal among fragments is restricted. Less frequent dispersal will result in relatively independent metapopulation dynamics with regular extinction and subsequent recolonization (Hanski, 1991). Severe fragmentation will lead to isolated like islands, where local extinctions will no longer be counterbalanced by colonization, and the possibility of dispersal depends on the distance between populations and the species' dispersal ability (Cheptou et al., 2017).

At the genetic level, small populations are more prone to genetic drift and inbreeding (A. Young et al., 1996). Loss of genetic diversity caused by drift shrinks potential to deal with novel environments. Simultaneously, deleterious mutations will tend to accumulate because the selection is less effective in small populations (Lynch et al., 1995). These effects do not threaten populations in the short term. On the other hand, inbreeding depression can act swiftly and reduce fitness, especially after rapid fragmentation (Keller and Waller, 2002). Fitness reductions from drift, inbreeding depression and the Allee effect can further reduce population

size, which will create a potentially dangerous feedback loop known as an extinction vortex (Cheptou et al., 2017)

2.2.1 Species response to fragmentation

Fragmented populations may deal with gene flow reduction by individual selection under local conditions within the population or in dependation on the frequency of dispersal also in metapopulation-level selection (Cote et al., 2017). Local selection can be driven by abiotic factors (e.g. habitat heterogeneity, edge effects) or biotic factor (e.g. herbivores) to avoid inbreeding depression (Bengtsson, 1978) and/ or competition with relatives (Hamilton and May, 1977), an increase of dispersion among populations is desired. By contrast, fragmentation can also act against dispersal because of increased dispersal cost (e.g. investments in seed morphology to optimize dispersal) (Bonte et al., 2012). Such contradictory pressures can result in fluctuating selection, where dispersal is promoted after initial colonization and declines subsequently, or can help maintain a variable dispersal strategy (Cheptou et al., 2017).

Populations can also deal with fragmentation through reproduction strategies. Less complicated reproduction like self-compatibility and self-fertilization is often associated with small population size (Baker, 1955).

Another possibility is the usage of available edges of habitat. Specialist species are more negatively affected by fragmentation, and generalist species can colonize surrounding areas better (Adriaens et al., 2006; Andrén et al., 1997). Moreover, specialist species are also considered low-quality dispersers because their dispersal distance is often short (N. Li et al., 2015). Species classified in the IUCN Red List categories in the Czech Republic as critically endangered, endangered and vulnerable were significantly more specialized than those classified in the category near threatened. The least concerned species were significantly less specialized than species classified in any other IUCN category (Zelený and Chytrý, 2019). One of the highly protected European endemic species is *Gentianella praecox subsp. bohemica*. This species is a strictly biennial herb endemic to the Bohemian Massif. Most populations occur in the Czech Republic but extend to Bavaria (Germany), Upper and Lower Austria, and Poland. *Gentianella* Moench (Gentianaceae) is a highly diverse and taxonomically complicated genus due to seasonal dimorphism, introgression, and hybridization between closely related species (Greimler and Jang, 2007; Plenk et al., 2016; Winfield et al., 2003).

Substantial population size reduction was recorded during the 20th century, probably due to land-use intensification or abandonment of traditional land use, which led to the disintegration

of large habitats and fragmentation of original populations. For genetic characterization of remaining populations was necessary to develop a set of polymorphic markers because no useful sequence information was available (Paper 5). We used 454 sequencing for the initial stage of microsatellite development (Malausa et al., 2011). Using this sequencing technology, a time-consuming and expensive cloning approach previously used (e.g. Duval *et al.*, 2005) to obtain microsatellite regions could be avoided. The second part of microsatellites genotyping used a traditional AB sequencer with standard protocol for fragmentation analysis. In Paper 6, thanks to this methodology, we obtain 373 regions with microsatellite motive. We were able to find 12 regions suitable for genotyping for an endemic, endangered species *Gentianella praecox subsp. bohemica*. These optimized reactions were used in Paper 6 to genotype 1400 individuals of *Gentianella praecox subsp. bohemica* and characterize the genetic composition of remaining populations. These populations go through frequent, short-term population fluctuations. However, surprisingly significant reductions in population size do not result in immediate loss of genetic diversity and reduction of fitness within the populations. This is likely due to the seed bank of the species serving as a reservoir of the genetic diversity of the populations.

Possibility for restoration of genetic diversity in seed banks was studied on many species (e.g., (McCue and Holtsford, 1998; Mandák *et al.*, 2006; Pujol *et al.*, 2007; Mandak *et al.*, 2012; Kremp *et al.*, 2016; Schulz *et al.*, 2018), including invasive (Fennell et al., 2014), aridic (Hanin et al., 2013) or aquatic (Uesugi et al., 2007) species. They confirmed that genetic diversity in the seed bank could be very high, may exceed that in the aboveground, and counteract effects of drift and selection.

The second studied endangered species, *Minuartia smejkalii,* is endemic to the Czech Republic. Being an obligate serpentinophyte, *M. smejkalii*'s natural occurrences are highly restricted by the patchy distribution of serpentine soils. Known habitats of *M. smejkalii* have been dramatically reduced in size and fragmented since 1960–1970 by human actions. Thus, contrary to their typically fragmented but stable-over-time serpentine habitats, the habitat of *M. smejkalii* underwent recent destruction followed by a substantial population decline. We investigate the population structure of survived populations using the AFLP- derived method, nextRAD (Nextera-tagmented reductively-amplified DNA) sequencing (Paper 7). This approach and its use and advantages for conservation studies were published by (Russello et al., 2015). In the case of *Minuartia smejkalii,* 119 individuals were sequenced, 1143 polymorphic SNPs were found. Based on identified SNPs, high genetic diversity within *M. smejkalii* populations, low to moderate among-populations differentiation, and moderate regional differentiation was

found. A highly outcrossing mating system can explain these results, or it is an effect of historical evolutionary processes. Interestingly, 2–3% of the markers show differentiation patterns consistent with divergent selection, suggesting that some local adaptation might have occurred.

Similar results have also been observed for other rare, endangered, and endemic species *Petunia secreta*, a Brazilian endemic with a very similar distribution to *M. smejkalii*, showed genetic differentiation between two regions about 20 km apart, but not within regions where the populations were <0.2 km apart (Turchetto et al., 2016). On the other hand, (Honnay et al., 2007) studied the genetic structure of 27 populations of the rare perennial plant species *Globularia bisnagarica* in Belgium. Given the mixed breeding system of the species and the relatively small area of the studied system, the populations were characterized by high genetic differentiation, significant isolation by distance of the populations. Average within population genetic diversity was low and was negatively related to population isolation. This suggests a more extensive gene flow into less isolated populations. Population size and local habitat characteristics did not significantly influence population genetic diversity. Furches, Small and Furches, 2013 studied three endangered pitcher plant species (Sarracenia; Sarraceniaceae) and found that species' distributions may not predict current levels of genetic variation and structure. Contrary to expectations, the most narrow-ranging species (*S. alabamensis*) exhibited the greatest variation. While *S. oreophila* lacked variation in microsatellite regions, it exhibited substantial chloroplast haplotype variation. Lastly, *Sarracenia jonesii* had the lowest overall genetic variation, lacks chloroplast diversity, yet maintains greater microsatellite diversity than *S. oreophila*. Authors conclude that three endangered species differ in levels and structure of diversity, yet not in predictable ways. Unique demographic and ecological histories, rather than current distribution and population size, best explain present patterns of genetic variation. These examples showed that response to fragmentation is time- species- and habitat- dependent.

The third species, *Rumex alpinus,* can be considered more as generalist compared to mentioned endemic species. This species naturally occur in isolated habitats in the French and Austrian Alps, in nutrient-rich habitats such as abandoned mountain meadows and along riverbanks. However, in the Czech Republic is it considered an invasive species. Although the species is known to be highly clonal (Klimeš, 1992), it is also known to spread both clonally and by seeds (Červenková and Münzbergová, 2009). However, the importance of these two modes of reproduction and their importance for the migration potential of the species was unknown. From identified 319 regions, 15 variable regions were suitable for multiplex PCR (Paper 8). We

genotyped 72 individuals from 6 isolated populations from the native species range, and we identified 1 to 4 alleles and observed and expected heterozygosities ranging from 0.06 to 1.0 and from 0.18 to 0.72, respectively. Significant deviation from Hardy–Weinberg equilibrium was found mainly in the native Crozet and Katschberg populations, indicating strong clonality. This indication is also supported for the Crozet population by lower observed heterozygosity than expected for each locus of this population. We additionally characterized 110 individuals from seven populations originating along riverbanks in the Czech Republic. However, they showed an extremely strong clonal effect, and we found no differentiation between these populations (unpublished data).

In line with our findings, clonal propagation in alpine species was repeatedly confirmed (Fischer *et al.*, 2000; Young *et al.*, 2002; Pluess and Stöcklin, 2004; De Witte *et al.*, 2012; Gaudeul, Delahaye and Muller, 2019; Varsamis *et al.*, 2021). Despite the fact, most clonal plants maintain sexual reproduction (Richards and Holsinger, 1997) and their populations exhibit considerable levels of genetic diversity (Ellstrand and Roose, 1987; Parker and Hamrick, 1992; Widén et al., 1994), clonal reproduction offers a solution for the harsh alpine environment. During life, alpine plants have to cope with irradiance, high UV levels, vapour pressure deficit, low temperatures, presence of snow and limited nutrient resources. This brings a high degree of uncertainty for their reproduction and wide-spreading. Clonal reproduction strategy allows adaptation to heterogeneous environments by risk spreading among genets, clonal integration, foraging behavior, and division of labor (Wang et al., 2018) and consequently successfully reproduce.

2.3 Overview of my research

Study 1: Selection on meiosis genes in diploid and tetraploid *Arabidopsis arenosa*.

This study compares 70 meiotic related genes in European diploid and tetraploid *Arabidopsis arenosa* populations. These genes were firstly characterized by whole genome sequencing and, in addition, also by long-range PCR. From 70 investigated genes, seven candidates genes were found to be strongly differentiated between populations. All seven genes encode structural proteins that coordinate processes in early prophase, including sister chromatid cohesion, axis formation and crossing over. Also, four of them showed outlier levels of differentiation between diploids and tetraploids, but not among diploids. In the following steps, high-frequency amino acid changes were examined. From 113 changes across subsection of the five genes and three population groups in *A. arenosa*, only in tetraploid populations, seven changes affect strongly conserved amino acids. Possible explanations for all presented results in this paper is that diploid and tetraploid *A. arenosa* experienced selection on structural proteins but probably from different background reasons.

Study 2: Unmasking cryptic biodiversity in polyploids: origin and diversification of Aster amellus aggregate.

Here we study the aggregate of *Aster amellus*, with identified diploid and hexaploid cytotypes. Although the cytotypes of *A. amellus* are not morphologically distinguishable, previous studies showed spatial segregation and limited gene flow between them, which could result in different evolutionary trajectories for each cytotype. We reconstructed the intraspecific-level phylogeny by sequencing internal transcribed spacer (ITS) region and non-coding chloroplast (pDNA) markers, and we combined Bayesian methods, population genetic analyses, statistical parsimony and ecological niche modelling tools to clarify the phylogeographic distribution patterns.

Nuclear and plastid Bayesian phylogenetic reconstructions showed different topologies. Analyses of plastid regions showed a topology where both cytotypes were shuffled within different clades. The nuclear phylogeny clearly separated the hexaploid individuals into one monophyletic clade and the diploid individuals into another monophyletic clade, except for one hexaploid individual. Haplotype network revealed nine plastid DNA haplotypes and seven different nuclear haplotypes. The plastid network showed that the two dominant haplotypes grouped populations, including both cytotypes. In the nuclear network, each haplotype was composed exclusively. Cytotypes occurring sympatrically in the mixed-ploidy population were

also pooled into different haplotypes in the network. Four haplotypes dominated and were widespread geographically: two of the diploid cytotype and two of the hexaploid cytotype. Ecological niche modelling showed that hexaploids hold a greater potential to occupy new areas to the east and partly also to the west of their current distribution.

Study 3: Absence of gene flow between diploids and hexaploids of *Aster amellus* at multiple spatial scales.

In this study, we characterized relationships between the diploid and hexaploid cytotypes of *Aster amellus* using seven highly variable microsatellite markers at three spatial scales: (1) within the single known mixed-ploidy population; (2) between pairs of diploid and hexaploid populations at the contact zone; and (3) between diploid and hexaploid individuals distributed across Europe. The results show a clear separation of the cytotypes at all three spatial scales. Furthermore, the AMOVA analyses for populations across contact zone, comparing the single-origin polyploidy model and strict multiple origins polyploidy model, clearly indicated that the single-origin model is superior to the multiple origin model. In the single-origin model, cytotype explained 26% of the total genetic variation and was highly significant. Some hexaploid individuals found in the mixed-ploidy population, however, fell into the diploid cluster. This could suggest recurrent polyploid formation or occasional cross-pollination between cytotypes. Moreover, as also suggested in Paper 2, this study propose that both cytotypes could represent cryptic species.

Study 4: Temperature and precipitation, but not geographic distance, explain genetic relatedness among populations in the perennial grass *Festuca rubra*.

In this study we asked how moisture and temperature influence genetic diversity of *Festuca rubra* populations and if genetic diversity is influenced by geographical position of populations. High levels of heterozygosity demonstrate that only the most heterozygous individuals were able to survive in the coldest and wettest region. Genotypic diversity decreased with increasing precipitation, and the decrease was stronger in the colder regions. Low genotypic diversity likely reflects extensive clonal growth, which is in line with the fact that the wet environment represents stressful conditions for the species. A high level of genetic diversity within populations is reflected in the absence of any isolation by distance for the total data set. This is likely because our study area shows extreme topographic variation, so geographic distance may limitedly reflect the connectivity among the populations. Also, variation within populations was significantly higher than the variation among populations, which correspond with species

dispersal abilities and life span. Comparison of the populations within the single temperature and precipitation levels demonstrated that the highest differentiation could be found in populations within the coldest and the wettest zone. It suggests strong selection pressure, which leads to strong between populations differentiation. In summary, both abiotic factors represent likely determinants of genetic relatedness in the system and the level of clonal growth of the species.

Study 5: Development of SSR markers by 454 sequencing in the endemic species *Gentianella praecox subsp. bohemica* (Gentianaceae).

This is a methodological study that describes the characterization of new microsatellite regions for assessing *Gentianella praecox subsp. bohemica* remaining populations. To identify microsatellite regions, 454 sequencing was used, and for efficient genotyping multiplex PCR was designed. As a result, we optimized 3 multiplex PCR mixes with 12 polymorphic regions.

Study 6: Strong fluctuations in aboveground population size do not limit genetic diversity in populations of an endangered biennial species.

In this study, we used optimized multiplex PCR reactions to evaluate the genetic diversity of *Gentianella praecox subsp. bohemica* remaining populations. Their population size was recorded last 15 years. Maximum population size over the last 15 years was the best predictor of expected heterozygosity of the populations and was significantly related to current population size and management. Plant fitness was significantly related to current as well as maximum population size and expected heterozygosity. Despite strong fluctuation in population size, the genetic diversity of this species is preserved.

Study 7: Low genetic differentiation despite high fragmentation in the endemic serpentinophyte *Minuartia smejkalii* (*M. verna* agg. Caryophyllaceae) revealed by RADSeq SNP markers.

In this study, we used nextRAD sequencing for revealing the genetic structure of remaining *Minuartia smejkalii* populations and closely related populations of *Minuartia corcontica* and *Minuartia caespitosa*. We found differentiation between all studied species. The AMOVA results showed that most of the genetic variation was found within populations. About 20% of the genetic variation was explained by the species. Also, 2–3% of the markers show differentiation patterns consistent with divergent selection, suggesting that some local adaptation might have occurred in *M. smejkalii*.

Study 8: Microsatellite markers for study of the invasive species *Rumex alpinus* (Polygonaceae).

This is a second methodological study describing microsatellite markers' development for population characterization of highly invasive species *Rumex alpinus*. To identify microsatellite regions, 454 sequencing was used, and 15 polymorphic regions were tested in European populations and proposed for population genotyping. We genotyped 72 individuals from 6 isolated populations from the native species range, and we identified 1 to 4 alleles and observed and expected heterozygosities ranging from 0.06 to 1.0 and from 0.18 to 0.72, respectively. Significant deviation from Hardy–Weinberg equilibrium was found mainly in the native Crozet and Katschberg populations, indicating strong clonality. Clonal growth is suggested as a preferred reproduction strategy and contributes greatly to the dispersal ability of this species.

2.4 Author contribution statement

1, Wright, K. M., Arnold, B., Xue, K., Šurinová, M., O'Connell, J., Bomblies, K. (2015). Selection on meiosis genes in diploid and tetraploid *Arabidopsis arenosa*. *Molecular biology and evolution*, 32(4), 944-955.

KB designed the experiments, KB and BA collected the plant material, KX, KW and MŠ performed the experiments and partially analyzed data, JC analyzed data, KW and KB wrote the manuscript; other authors provided editorial advice.

2, Mairal, M., Šurinová, M., Castro, S., and Münzbergová, Z. (2018). Unmasking cryptic biodiversity in polyploids: origin and diversification of *Aster amellus* aggregate. *Annals of Botany*, 122(6), 1047-1059.

ZM and SC conceived and designed the experiments, SC collected the plant material, MŠ performed the experiments and partially analyzed data, MM completed data analyses and wrote the manuscript, ZM revised the manuscript; other authors provided editorial advice.

3, Münzbergová, Z., Šurinová, M., and Castro, S. (2013). Absence of gene flow between diploids and hexaploids of *Aster amellus* at multiple spatial scales. *Heredity*, 110(2), 123-130.

ZM and SC designed the experiments, SC collected the plant material, MŠ performed the experiments and partially analyzed data, ZM completed data analyses and wrote the manuscript; other authors provided editorial advice.

4, Šurinová, M., Hadincová, V., Vandvik, V., and Münzbergová, Z. (2019). Temperature and precipitation, but not geographic distance, explain genetic relatedness among populations in the perennial grass *Festuca rubra*. *Journal of Plant Ecology*, 12(4), 730-741.

All authors designed the experiments, VH and ZM collected the plant material, MŠ performed the experiments and analyzed data, MŠ and ZM jointly wrote the manuscript; other authors provided editorial advice.

5, Šurinová, M., Brabec, J., and Münzbergová, Z. (2017). Development of SSR markers by 454 sequencing in the endemic species *Gentianella praecox subsp. bohemica* (Gentianaceae). *Applications in plant sciences*, 5(1), 1600114.

MŠ designed the experiments, JB collected the plant material, MŠ performed the experiments and analyzed data, MŠ wrote the manuscript, ZM revised the manuscript.

6, Münzbergová, Z., Šurinová, M., Husáková, I., and Brabec, J. (2018). Strong fluctuations in aboveground population size do not limit genetic diversity in populations of an endangered biennial species. *Oecologia*, 187(3), 863-872.

ZM and JB conceived and designed the experiments, JB and IH collected the plant material, MŠ performed the experiments, MŠ and ZM analyzed the data, IH created the maps, ZM wrote the manuscript; other authors provided editorial advice.

7, Stojanova, B., Šurinová, M., Zeisek, V., Münzbergová, Z., Pánková, H. (2020). Low genetic differentiation despite high fragmentation in the endemic serpentinophyte *Minuartia smejkalii* (M. verna agg. Caryophyllaceae) revealed by RADSeq SNP markers. *Conservation Genetics*, 21(2), 187-198.

ZM and HP designed the experiments, HP collected the plant material, MŠ designed laboratory experiments, VZ and BS analyzed data, BS, MŠ, VZ, HP jointly wrote the manuscript; ZM revised the manuscript.

8, Šurinová, M., Jarošová, A., and Münzbergová, Z. (2018). Microsatellite markers for study of the invasive species *Rumex alpinus* (Polygonaceae). *Applications in plant sciences*, 6(1), e1017.

MŠ designed the experiments, AJ and MŠ performed the experiments, MŠ analyzed data and wrote the manuscript, ZM revised the manuscript.

Prof. Zuzana Münzbergová Ph.D.

2.5 References

Adriaens, D., Honnay, O., Hermy, M. 2006. No evidence of a plant extinction debt in highly fragmented calcareous grasslands in Belgium. Biological Conservation 133:212–224.

Ahrens, C. W., James, E. A., Miller, A. D., Scott, F., Aitken, N. C., Jones, A. W., Lu-Irving, P., Borevitz, J. O., Cantrill, D. J., Rymer, P. D. 2020. Spatial, climate and ploidy factors drive genomic diversity and resilience in the widespread grass *Themeda triandra*. Molecular Ecology 29:3872–3888.

Andrén, H., Delin, A., Seiler, A. 1997. Population response to landscape changes depends on specialization to different landscape elements. Oikos 80:193–196.

Arrigo, N., de La Harpe, M., Litsios, G., Zozomová-Lihová, J., Španiel, S., Marhold, K., Barker, M. S., Alvarez, N. 2016. Is hybridization driving the evolution of climatic niche in *Alyssum montanum*. American Journal of Botany 103:1348–1357.

Baker, H. G. 1955. Self-compatibility and establishment after'long-distance'dispersal. Evolution 9:347–349.

Bengtsson, B. O. 1978. Avoiding inbreeding: at what cost? Journal of Theoretical Biology 73:439–444.

Blaine Marchant, D., Soltis, D. E., Soltis, P. S. 2016. Patterns of abiotic niche shifts in allopolyploids relative to their progenitors. New Phytologist 212:708–718.

Bonte, D., Van Dyck, H., Bullock, J. M., Coulon, A., Delgado, M., Gibbs, M., Lehouck, V., Matthysen, E., Mustin, K., Saastamoinen, M. 2012. Costs of dispersal. Biological Reviews 87:290–312.

Brennan, A. C., Bridle, J. R., Wang, A., Hiscock, S. J., Abbott, R. J. 2009. Adaptation and selection in the Senecio (Asteraceae) hybrid zone on Mount Etna, Sicily. New Phytologist 183:702–717.

Brito, V. L. G., Mori, G. M., Vigna, B. B. Z., Azevedo-Silva, M., Souza, A. P., Sazima, M. 2016. Genetic structure and diversity of populations of polyploid *Tibouchina pulchra* Cogn.(Melastomataceae) under different environmental conditions in extremes of an elevational gradient. Tree Genetics and Genomes 12:1–12.

Brownfield, L., and Köhler, C. 2011. Unreduced gamete formation in plants: mechanisms and

prospects. Journal of Experimental Botany 62:1659–1668.

Čertner, M., Fenclová, E., Kúr, P., Kolár, F., Koutecký, P., Krahulcová, A., Suda, J. 2017. Evolutionary dynamics of mixed-ploidy populations in an annual herb: dispersal, local persistence and recurrent origins of polyploids. Annals of Botany 120:303–315.

Červenková, Z. , and Münzbergová Z.. 2009. Susceptibility of the landscape of the Giant Mountains, Czech Republic, to invasion by *Rumex alpinus* In Pyšek P. and Pergl J. [eds.], Biological invasions: Towards a synthesis. Neobiota 8:75–85.

Chen, Z. J., and Ni, Z. 2006. Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. Bioessays 28:240–252.

Cheptou, P.O., Hargreaves, A. L., Bonte, D., Jacquemyn, H. 2017. Adaptation to fragmentation: evolutionary dynamics driven by human influences. Philosophical Transactions of the Royal Society B: Biological Sciences 372:20160037.

Collins, C. D., Holt, R. D., Foster, B. L. 2009. Patch size effects on plant species decline in an experimentally fragmented landscape. Ecology 90:2577–2588.

Cote, J., Bestion, E., Jacob, S., Travis, J., Legrand, D., Baguette, M. 2017. Evolution of dispersal strategies and dispersal syndromes in fragmented landscapes. Ecography 40:56–73.

De Witte, L. C., Armbruster, G. F. J., Gielly, L., Taberlet, P., Stocklin, J. 2012. AFLP markers reveal high clonal diversity and extreme longevity in four key arctic-alpine species. Molecular Ecology 21:1081–1097.

Duchoslav, M., Jandová, M., Kobrlová, L., Šafářová, L., Brus, J., Vojtěchová, K. 2020. Intricate distribution patterns of six cytotypes of *Allium oleraceum* at a continental scale: niche expansion and innovation followed by niche contraction with increasing ploidy level. Frontiers in Plant Science 11:1885-1908.

Duval, M.F., Bunel, J., Sitbon, C., Risterucci, A.M. 2005. Development of microsatellite markers for mango (*Mangifera indica* L.). Molecular Ecology Notes 5:824–826.

Ellstrand, N. C., Roose, M. L. 1987. Patterns of genotypic diversity in clonal plant species. American Journal of Botany 74:123–131.

Eriksson, J. S., Blanco-Pastor, J. L., Sousa, F., Bertrand, Y. J. K., Pfeil, B. E. 2017. A cryptic

species produced by autopolyploidy and subsequent introgression involving *Medicago prostrata* (Fabaceae). Molecular Phylogenetics and Evolution 107:367–381.

Ewers, R. M., and Didham, R. K. 2006. Confounding factors in the detection of species responses to habitat fragmentation. Biological Reviews 81:117–142.

Fennell, M., Gallagher, T., Vintro, L. L., Osborne, B. 2014. Using soil seed banks to assess temporal patterns of genetic variation in invasive plant populations. Ecology and Evolution 4:1648–1658.

Fischer, M., Husi, R., Prati, D., Peintinger, M., van Kleunen, M., Schmid, B. 2000. RAPD variation among and within small and large populations of the rare clonal plant *Ranunculus reptans* (Ranunculaceae). American Journal of Botany 87:1128–1137.

Freeland, J. R., Petersen, S. D., Kirk, H. 2011. Molecular ecology. Wiley-Blackwell.

Furches, M. S., Small, R. L., Furches, A. 2013. Genetic diversity in three endangered pitcher plant species (Sarracenia; Sarraceniaceae) is lower than widespread congeners. American Journal of Botany 100:2092–2101.

Gaudeul, M., Delahaye, T., Muller, S. 2019. AFLP markers show low levels of clonal propagation and high genotypic diversity in the rare, southernmost populations of *Linnaea borealis* L. (Caprifoliaceae) in the Western Alps. Genetica 147:79–90.

Godfree, R. C., Marshall, D. J., Young, A. G., Miller, C. H., Mathews, S. 2017. Empirical evidence of fixed and homeostatic patterns of polyploid advantage in a keystone grass exposed to drought and heat stress. Royal Society Open Science 4:170934.

Goga, M., Ručová, D., Kolarčik, V., Sabovljević, M., Bačkor, M., Lang, I. 2018. Usnic acid, as a biotic factor, changes the ploidy level in mosses. Ecology and Evolution 8:2781–2787.

Greimler, J., and Jang, C.G. 2007. Gentianella stiriaca, a case of reticulate evolution in the northeastern and eastern Central Alps. Taxon 56:857–870.

Groom, M. J., and Schumaker, N. 1993. Evaluating landscape change: patterns of worldwide deforestation and local fragmentation. Biotic Interactions and Global Change 24–44.

Hamilton, W. D., and May, R. M. 1977. Dispersal in stable habitats. Nature 269:578–581.

Hanin, N., Quaye, M., Westberg, E., Barazani, O. 2013. Soil seed bank and among-years

genetic diversity in arid populations of *Eruca sativa* Miller (Brassicaceae). Journal of Arid Environments 91:151–154.

Hanski, I. 1991. Single-species metapopulation dynamics: concepts, models and observations. Biological Journal of the Linnean Society 42:17–38.

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

Hollister, J. D. 2015. Polyploidy: adaptation to the genomic environment. New Phytologist 205:1034–1039.

Honnay, O., Adriaens, D., Coart, E., Jacquemyn, H., Roldan-Ruiz, I. 2007. Genetic diversity within and between remnant populations of the endangered calcareous grassland plant *Globularia bisnagarica* L. Conservation Genetics 8:293–303.

Hulshof, C. M., and Spasojevic, M. J. 2020. The edaphic control of plant diversity. Global Ecology and Biogeography 29:1634–1650.

Husband And, B. C., and Schemske, D. W. 2000. Ecological mechanisms of reproductive isolation between diploid and tetraploid *Chamerion angustifolium*. Journal of Ecology 88:689–701.

Karunarathne, P., Schedler, M., Martínez, E. J., Honfi, A. I., Novichkova, A., Hojsgaard, D. 2018. Intraspecific ecological niche divergence and reproductive shifts foster cytotype displacement and provide ecological opportunity to polyploids. Annals of Botany 121:1183–1196.

Keller, L. F., and Waller, D. M. 2002. Inbreeding effects in wild populations. Trends in Ecology and Evolution 17:230–241.

Kiedrzyński, M., Zielińska, K. M., Jedrzejczyk, I., Kiedrzyńska, E., Tomczyk, P. P., Rewicz, A., Rewers, M., Indreica, A., Bednarska, I., Stupar, V. 2021. Tetraploids expanded beyond the mountain niche of their diploid ancestors in the mixed-ploidy grass *Festuca amethystina* L. Scientific Reports 11:1–12.

Klimeš, L. 1992. The clone architecture of *Rumex alpinus* (Polygonaceae). Oikos 402–409.

Kotliar, N. B., and Wiens, J. A. 1990. Multiple scales of patchiness and patch structure: a hierarchical framework for the study of heterogeneity. Oikos 253–260.

Kremp, A., Oja, J., LeTortorec, A. H., Hakanen, P., Tahvanainen, P., Tuimala, J., Suikkanen, S. 2016. Diverse seed banks favour adaptation of microalgal populations to future climate conditions. Environmental Microbiology 18:679–691.

Laport, R. G., Minckley, R. L., Ramsey, J. 2016. Ecological distributions, phenological isolation, and genetic structure in sympatric and parapatric populations of the *Larrea tridentata* polyploid complex. American Journal of Botany 103:1358–1374.

Laurance, W. F., Lovejoy, T. E., Vasconcelos, H. L., Bruna, E. M., Didham, R. K., Stouffer, P. C., Gascon, C., Bierregaard, R. O., Laurance, S. G., Sampaio, E. 2002. Ecosystem decay of Amazonian forest fragments: a 22-year investigation. Conservation Biology 16:605–618.

Li, N., Fang, S., Li, X., An, S., Lu, C. 2015. Differential contribution of frugivorous birds to dispersal patterns of the endangered Chinese yew (*Taxus chinensis*). Scientific Reports 5:1–7.

Li, W., Berlyn, G. P., Ashton, P. M. S. 1996. Polyploids and their structural and physiological characteristics relative to water deficit in *Betula papyrifera* (Betulaceae). American Journal of Botany 83:15–20.

Li, X., and Guo, B. 2020. Substantially adaptive potential in polyploid cyprinid fishes: evidence from biogeographic, phylogenetic and genomic studies. Proceedings of the Royal Society B 287:20193008.

Lynch, M., Conery, J., Burger, R. 1995. Mutation accumulation and the extinction of small populations. The American Naturalist 146: 489–518.

Maherali, H., Walden, A. E., Husband, B. C. 2009. Genome duplication and the evolution of physiological responses to water stress. New Phytologist 184:721–731.

Malausa, T., Gilles, A., Meglécz, E., Blanquart, H., Duthoy, S., Costedoat, C., Dubut, V., Pech, N., Castagnone-Sereno, P., Delye, C. 2011. High-throughput microsatellite isolation through 454 GS-FLX Titanium pyrosequencing of enriched DNA libraries. Molecular Ecology Resources 11:638–644.

Mandák, B., Bímová, K., Mahelka, V., Plačková, I. 2006. How much genetic variation is stored in the seed bank? A study of *Atriplex tatarica* (Chenopodiaceae). Molecular Ecology 15:2653–2663.

Mandak, B., Zákravský, P., Mahelka, V., Plačková, I. 2012. Can soil seed banks serve as genetic memory? A study of three species with contrasting life history strategies. PLoS One 7:e49471.

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

Manzaneda, A. J., Rey, P. J., Anderson, J. T., Raskin, E., Weiss-Lehman, C., Mitchell-Olds, T. 2015. Natural variation, differentiation, and genetic trade-offs of ecophysiological traits in response to water limitation in *Brachypodium distachyon* and its descendent allotetraploid *B. hybridum* (Poaceae). Evolution 69:2689–2704.

Manzaneda, A. J., Rey, P. J., Bastida, J. M., Weiss-Lehman, C., Raskin, E., Mitchell-Olds, T. 2012. Environmental aridity is associated with cytotype segregation and polyploidy occurrence in *Brachypodium distachyon* (Poaceae). New Phytologist 193:797–805.

McCue, K. A., and Holtsford, T. P. 1998. Seed bank influences on genetic diversity in the rare annual *Clarkia springvillensis* (Onagraceae). American Journal of Botany 85:30–36.

Melaragno, J. E., Mehrotra, B., Coleman, A. W. 1993. Relationship between endopolyploidy and cell size in epidermal tissue of *Arabidopsis*. The Plant Cell 5:1661–1668.

Nassar, N. M. A., Graciano-Ribeiro, D., Fernandes, S. D. C., Araujo, P. C. 2008. Anatomical alterations due to polyploidy in cassava, *Manihot esculenta* Crantz. Genetics and Molecular Research 7:276–283.

Otto, S. P., and Michalakis, Y. 1998. The evolution of recombination in changing environments. Trends in Ecology and Evolution 13:145–151.

Page, S. L., and Hawley, R. S. 2003. Chromosome choreography: the meiotic ballet. Science 301:785–789.

Parker, K. C., and Hamrick, J. L. 1992. Genetic diversity and clonal structure in a columnar cactus, *Lophocereus schottii*. American Journal of Botany 79:86–96.

Plenk, K., Göd, F., Kriechbaum, M., Kropf, M. 2016. Genetic and reproductive characterisation of seasonal flowering morphs of *Gentianella bohemica* revealed strong reproductive isolation and possible single origin. Plant Biology 18:111–123.

Pluess, A. R., and Stöcklin, J. 2004. Population genetic diversity of the clonal plant *Geum*

reptans (Rosaceae) in the Swiss Alps. American Journal of Botany 91:2013–2021.

Pujol, B., Renoux, F., Elias, M., Rival, L., McKey, D. 2007. The unappreciated ecology of landrace populations: Conservation consequences of soil seed banks in Cassava. Biological Conservation 136:541–551.

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

Ramsey, J., and Schemske, D. W. 2002. Neopolyploidy in flowering plants. Annual Review of Ecology and Systematics 33:589–639.

Rice, A., Šmarda, P., Novosolov, M., Drori, M., Glick, L., Sabath, N., Meiri, S., Belmaker, J., Mayrose, I. 2019. The global biogeography of polyploid plants. Nature Ecology and Evolution 3:265–273.

Richards, A. J., and Holsinger, K. E. 1997. Plant breeding systems (2nd edn). Trends in Plant Science 2: 485.

Rodriguez, D. J. 1996. A model for the establishment of polyploidy in plants. The American Naturalist 147:33–46.

Russello, M. A., Waterhouse, M. D., Etter, P. D., Johnson, E. A. 2015. From promise to practice: pairing non-invasive sampling with genomics in conservation. PeerJ 3:e1106–e1106.

Rybicki, J., and Hanski, I. 2013. Species–area relationships and extinctions caused by habitat loss and fragmentation. Ecology Letters 16:27–38.

Schulz, B., Durka, W., Danihelka, J., Eckstein, R. L. 2018. Differential role of a persistent seed bank for genetic variation in early vs. late successional stages. PLoS One 13:e0209840.

Seear, P. J., France, M. G., Gregory, C. L., Heavens, D., Schmickl, R., Yant, L., Higgins, J. D. 2020. A novel allele of ASY3 is associated with greater meiotic stability in autotetraploid *Arabidopsis lyrata*. PLOS Genetics 16:e1008900.

Shaffer, M. L. 1981. Minimum population sizes for species conservation. BioScience 31:131–134.

Silvertown, J., and Charlesworth, D. 2001. Introduction to plant population biology.

Blackwell Science, Oxford.

Stephens, P. A., Sutherland, W. J., Freckleton, R. P. 1999. What Is the Allee Effect? Oikos 87:185–190.

Stevens, A. V, Nicotra, A. B., Godfree, R. C., Guja, L. K. 2020. Polyploidy affects the seed, dormancy and seedling characteristics of a perennial grass, conferring an advantage in stressful climates. Plant Biology 22:500–513.

Sullivan, E. R., Barker, C., Powell, I., Ashton, P. A. 2019. Genetic diversity and connectivity in fragmented populations of *Rhinanthus minor* in two regions with contrasting land-use. Biodiversity and Conservation 28:3159–3181.

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

Turchetto, C., Segatto, A. L. A., Mäder, G., Rodrigues, D. M., Bonatto, S. L., Freitas, L. B. 2016. High levels of genetic diversity and population structure in an endemic and rare species: implications for conservation. AoB Plants 8.

Uesugi, R., Nishihiro, J., Tsumura, Y., Washitani, I. 2007. Restoration of genetic diversity from soil seed banks in a threatened aquatic plant, *Nymphoides peltata*. Conservation Genetics 8:111–121.

Varsamis, G., Merou, T., Karapatzak, E., Papageorgiou, A. C., Fotiadis, G., Tsiftsis, S. 2021. Genetic diversity of alpine *Dryas octopetala* populations at their southern distribution limit in Europe. Nordic Journal of Botany 39: e03150.

Wallace, J. B., Eggert, S. L., Meyer, J. L., Webster, J. R. 1997. Multiple trophic levels of a forest stream linked to terrestrial litter inputs. Science 277:102–104.

Wang, X., Zhao, W., Li, L., You, J., Ni, B., Chen, X. 2018. Clonal plasticity and diversity facilitates the adaptation of *Rhododendron aureum* Georgi to alpine environment. PLoS One 13:e0197089.

Weiss, H., Dobeš, C., Schneeweiss, G. M., Greimler, J. 2002. Occurrence of tetraploid and hexaploid cytotypes between and within populations in Dianthus sect. Plumaria (Caryophyllaceae). New Phytologist 156:85–94.

Widén, B., Cronberg, N., Widén, M. 1994. Genotypic diversity, molecular markers and spatial distribution of genets in clonal plants, a literature survey. Folia Geobotanica 29:245–263.

Winfield, M. O., Wilson, P. J., Labra, M., Parker, J. S. 2003. A brief evolutionary excursion comes to an end: the genetic relationship of British species of Gentianella sect. Gentianella (Gentianaceae). Plant Systematics and Evolution 237:137–151.

Yamauchi, A., Hosokawa, A., Nagata, H., Shimoda, M. 2004. Triploid bridge and role of parthenogenesis in the evolution of autopolyploidy. The American Naturalist 164:101–112.

Young, A., Boyle, T., Brown, T. 1996. The population genetic consequences of habitat fragmentation for plants. Trends in Ecology and Evolution 11:413–418.

Young, A. G., Hill, J. H., Murray, B. G., Peakall, R. 2002. Breeding system, genetic diversity and clonal structure in the sub-alpine forb *Rutidosis leiolepis* F. Muell. (Asteraceae). Biological Conservation 106:71–78.

Zelený, D., and Chytrý, M. 2019. Ecological specialization indices for species of the Czech flora. Preslia 91:93–116.

Zhang, J., Wang, M., Guo, Z., Guan, Y., Guo, Y., Yan, X. 2019. Variation in ploidy level and genome size of *Cynodon dactylon* (L.) Pers. along a latitudinal gradient. Folia Geobotanica 54:267–278.

Zozomová-Lihová, J., Malánová-Krásná, I., Vít, P., Urfus, T., Senko, D., Svitok, M., Kempa, M., Marhold, K. 2015. Cytotype distribution patterns, ecological differentiation, and genetic structure in a diploid–tetraploid contact zone of *Cardamine amara*. American Journal of Botany 102:1380–1395.

Study 1: Selection on meiosis genes in diploid and tetraploid *Arabidopsis arenosa*

Selection on Meiosis Genes in Diploid and Tetraploid *Arabidopsis arenosa*

Kevin M. Wright,[1] Brian Arnold,[1] Katherine Xue,[1] Maria Šurinová,[2] Jeremy O'Connell,[1] and Kirsten Bomblies*[,1]

[1]Department of Evolutionary and Organismic Biology, Harvard University

[2]Institute of Botany, Academy of Sciences of the Czech Republic, Pruhonice, Czech Republic

***Corresponding author:** E-mail: kbomblies@oeb.harvard.edu.

Associate editor: Stephen Wright

Abstract

Meiotic chromosome segregation is critical for fertility across eukaryotes, and core meiotic processes are well conserved even between kingdoms. Nevertheless, recent work in animals has shown that at least some meiosis genes are highly diverse or strongly differentiated among populations. What drives this remains largely unknown. We previously showed that autotetraploid *Arabidopsis arenosa* evolved stable meiosis, likely through reduced crossover rates, and that associated with this there is strong evidence for selection in a subset of meiosis genes known to affect axis formation, synapsis, and crossover frequency. Here, we use genome-wide data to study the molecular evolution of 70 meiosis genes in a much wider sample of *A. arenosa*. We sample the polyploid lineage, a diploid lineage from the Carpathian Mountains, and a more distantly related diploid lineage from the adjacent, but biogeographically distinct Pannonian Basin. We find that not only did selection act on meiosis genes in the polyploid lineage but also independently on a smaller subset of meiosis genes in Pannonian diploids. Functionally related genes are targeted by selection in these distinct contexts, and in two cases, independent sweeps occurred in the same loci. The tetraploid lineage has sustained selection on more genes, has more amino acid changes in each, and these more often affect conserved or potentially functional sites. We hypothesize that Pannonian diploid and tetraploid *A. arenosa* experienced selection on structural proteins that mediate sister chromatid cohesion, the formation of meiotic chromosome axes, and synapsis, likely for different underlying reasons.

Key words: **meiosis, evolution, polyploidy.**

Article

Introduction

Reliable meiotic recombination and chromosome segregation are essential for the persistence of virtually all eukaryotic species. Given that the basic features of meiosis are largely conserved across even wide evolutionary distances (Villeneuve and Hillers 2001; Gerton and Hawley 2005; Mercier and Grelon 2008; Harrison et al. 2010), we might expect that the proteins that coordinate key steps in meiosis should be highly constrained. However, recent studies in animals show that this is not necessarily the case—many meiosis genes, particularly those involved in the prophase I processes of chromosome pairing, DNA repair and homologous recombination, are often surprisingly divergent among populations, and in some cases even variable within them (e.g., Turner et al. 2008; Anderson et al. 2009; Chowdhury et al. 2009; Kong et al. 2014). For example, genes implicated in meiotic DNA repair, homologous recombination, and chromosome segregation show signatures suggestive of adaptive divergence among *Drosophila melanogaster* populations, as well as between *D. melanogaster* and *D. simulans* (Turner et al. 2008; Anderson et al. 2009). Four meiosis-related genes were among those differentiated between mosquito populations along a latitudinal cline (Cheng et al. 2012), and in humans a *STRUCTURAL MAINTENANCE OF CHROMOSOMES* (*SMC*) gene involved in sister chromatid cohesion and crossover (CO) regulation shows evidence of having undergone a selective sweep in Chinese populations (Williamson et al. 2007). Although the mechanisms driving adaptive evolution of proteins important for meiosis remain unclear, they can have direct or indirect pleiotropic effects on genome-wide recombination rates and fertility. For instance, variants in several meiosis-related genes in humans have been statistically associated with effects on recombination and fertility. These genes include *RAD21L*, which encodes part of the sister chromatid cohesion complex and is required for axis formation and synapsis, *MSH4* (*mutS homolog 4*), which encodes a ZMM protein required for CO formation, and *RNF212*, which encodes another ZMM protein essential for synapsis and recombination (Kong et al. 2008, 2014; Chowdhury et al. 2009; Fledel-Alon et al. 2011). In addition, variation in female genome-wide recombination rates is associated with an inversion, 17q21.31 (Fledel-Alon et al. 2011; Kong et al. 2014), which shows evidence of selection in Europeans (Stefansson et al. 2005; Boettger et al. 2012; Steinberg et al. 2012), though it is still unclear whether the structural variant itself, or a gene within this approximately 1-Mb region, is actually responsible for recombination rate variation. For other systems, it remains unknown whether divergence in meiosis genes is common, or whether this is causally linked with CO variation or other differences in meiosis.

One context in which we might expect selection to drive the evolution of meiosis genes is after whole-genome

© The Author 2014. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

duplication (WGD), which leads to polyploidy. WGD occurs in all eukaryotic kingdoms and is thought to have contributed to organismal complexity and genomic novelty (Ramsey and Schemske 1998; Otto and Whitton 2000; Soltis et al. 2003; Comai 2005; Rieseberg and Willis 2007; Doyle et al. 2008; Bomblies and Madlung 2014). However, when it initially occurs, WGD poses significant challenges to chromosome segregation in meiosis (e.g., Myers 1945; Gilles and Randolph 1951; Bremer and Bremer-Reinders 1954; Ramsey and Schemske 1998; Comai 2005). In some polyploids, meiotic stability evolves through an increase in pairing partner preferences, whereas in species that retain random segregation of all homologs, lower genome-wide CO rates can promote fertility by reducing the formation of multivalent associations among multiple homologs in meiosis (e.g., Hazarika and Rees 1967; Wolf et al. 1989; Khawaja et al. 1997; Cifuentes et al. 2010; Bomblies and Madlung 2014; Lavania 1986). Autotetraploid *Arabidopsis arenosa* fits this latter model, as it has diploid-like bivalent formation despite random chromosome pairing, coupled with a significantly lower genome-wide CO rate than its diploid relatives (Comai et al. 2003; Yant et al. 2013). Patterns of genetic differentiation between ploidy levels suggest that the tetraploid lineage experienced strong selection on genes involved in cohesion, axis formation, and synapsis, all of which are implicated in altering genome-wide CO rates (Hollister et al. 2012; Yant et al. 2013). Expression level of one of these genes, *ASY1*, has also been shown through transgenic work in tetraploid wheat to directly affect CO rates and multivalent formation (Boden et al. 2009).

Here, we sought to better understand the evolution of 70 meiosis genes compared with genome-wide patterns of nucleotide diversity in a much broader sampling of populations of both diploid and autotetraploid *A. arenosa*. This species is an obligately outcrossing relative of *A. thaliana* with high genetic diversity (Lind-Halldén et al. 2002; Hollister et al. 2012; Schmickl et al. 2012). There is evidence that gene flow among populations of the same ploidy is high, and the widespread autotetraploids comprise a single lineage (B.A. and K.B. unpublished data), with limited gene flow from the diploid to the tetraploid lineage occurring within the Carpathian Mountains (Jørgensen et al. 2011; Schmickl et al. 2012). We sampled six tetraploid populations from throughout the range, as well as six diploid populations from two adjacent, biogeographically distinct regions: The Carpathian Mountains and the Pannonian Basin. Diploids from the two regions comprise distinct genetic lineages, with Carpathian Mountain diploids more closely related to the tetraploids. We find evidence that selection acted on meiosis genes not only in the tetraploid lineage but also in the diploid populations, extending patterns seen previously in *Drosophila*, mosquitoes, and humans to plants. We show that amino acid substitutions with large differences in allele frequency only rarely affect strongly conserved sites, or sites predicted to be phosphorylated, sumoylated, or in functional domains; all seven exceptions are unique to the autotetraploid lineage. We also find that even though in each lineage multiple interacting genes are under selection, the derived polymorphisms are rare or absent in related populations. Our results, together with previous findings, paint a picture that a subset of meiosis genes may repeatedly experience selection, though perhaps for different reasons in different lineages.

Results

Sampling and Population Relatedness

To analyze the molecular evolution of meiosis genes in a greater swath of *A. arenosa* diversity than we sampled previously (Hollister et al. 2012; Yant et al. 2013), we whole genome sequenced pooled samples (PoolSeq) of 165 individuals from 12 populations, 6 of each ploidy (fig. 1*A* and supplementary table S1, Supplementary Material online). We aligned the sequences to the *A. lyrata* genome (Hu et al. 2011) using the

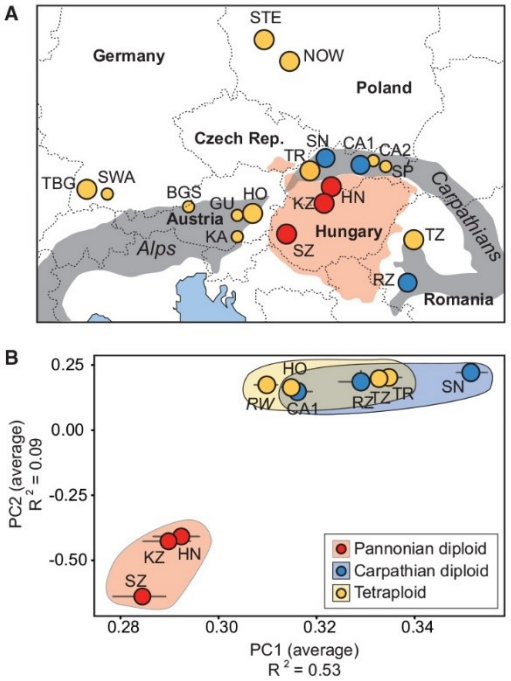


FIG. 1. Populations sampled in this study. (*A*) Population locations. Yellow circles indicate tetraploid populations sampled for PoolSeq data set (large circles), small circles are additional tetraploid populations. Blue circles indicate Carpathian diploids from the Carpathian Mountains. Red circles indicate Pannonian diploid populations from Hungary and Slovakia. Red-shaded area shows the extent of the Pannonian Basin biogeographic zone, and gray-shaded regions indicate alpine biogeographic zones of the Alps and Carpathians. (*B*) PCA of whole-genome PoolSeq of ten populations. Each circle is an average PC value for an entire population corresponding to the map in panel (*A*), with the exception of RW, which is a mixed pool of three genetically very similar railway populations, STE, NOW, and TBG. PCA was conducted using 290,000 informative sites present in at least seven of ten populations. To account for missing data in particular genomic regions, we randomly subsampled 50,000 sites 100 times, conducted independent PCAs, and presented the 95% confidence intervals as error bars for each population. GPS coordinates of all populations are given in supplementary table S1, Supplementary Material online.

program STAMPY (Lunter and Goodson 2011) and estimated allele frequency in each pool with SNAPE (Raineri et al. 2012). We analyzed the genetic relatedness of sampled populations using Principal Component Analysis (PCA) with 290,000 high confidence markers. Tetraploids are most closely related to diploids from the Carpathian Mountains (hereafter "Carpathian diploids") whereas samples from the Pannonian Basin of Slovakia and Hungary, which is a distinct biogeographic region (Molnár et al. 2006; Erdős et al. 2014), form a second more distantly related diploid gene pool (fig. 1B; hereafter "Pannonian diploids"). Using Worldclim data (Hijmans et al. 2005), we confirmed that the "Pannonian" sites we collected from are significantly warmer than sites where Carpathian diploids or tetraploids were collected (supplementary fig. S1, Supplementary Material online). Based on genetic relatedness and ploidy, we classified populations into each of three groups for subsequent analyses: Tetraploids (Tet), Carpathian diploids (C), and Pannonian diploids (P).

Differentiation of Meiosis Genes among Three *A. arenosa* Gene Pools

We tested which (if any) of 70 meiosis genes (Yant et al. 2013) show genome-wide outlier levels of differentiation among lineages using three different measurements. We used F_{ST} (Weir and Cockerham 1984), *G*, a statistic that quantifies raw allelic differentiation accounting for variation in read coverage (Magwene et al. 2011), and DD, which measures the ratio between the difference in allele frequency between populations and diversity within a population to account for the positive relationship between diversity and differentiation (Yant et al. 2013). We identified a total of 9.45 million single nucleotide polymorphisms (SNPs) for the Carpathian and Pannonian diploid contrasts (C × P), 8.71 million SNPs for the Pannonian diploid by tetraploid contrast (P × Tet), and 11.74 million SNPs for the Carpathian diploid by tetraploid contrast (C × Tet). We calculated each statistic for windows spanning 20, 40, or 100 SNPs for comparisons between all pairs of populations ($n = 45$) as well as between the three groups, and assigned rank scores to each window based on its position in the genome-wide distribution of values for F_{ST}, *G*, and DD (supplementary table S2, Supplementary Material online). Despite variation in coverage between population groups, this had almost no effect on our measurement of genetic diversity (supplementary table S3, Supplementary Material online). We next tested whether any of the 70 meiosis-related genes is enriched, compared with genome-wide background, for windows with outlier values of all three statistics. We identified seven genes (*ASY1, ASY3 PDS5, ZYP1a, ZYP1b, SMC1*, and *SYN1/AtREC8*) that each had multiple outlier windows for the three statistics in at least one set of populations (table 1). These results are largely consistent across window sizes (supplementary table S2, Supplementary Material online). For simplicity, we focus on one size class (40 SNP windows) in further analyses. Candidate genes show outlier values for each of the three statistics in comparisons between population pools, but in most cases, not among populations within pools (supplementary fig. S2, Supplementary Material online). All seven genes encode structural proteins that coordinate key events in early prophase including sister chromatid cohesion, axis formation, and crossing over (Hollingsworth et al. 1990; Bai et al. 1999; Panizza et al. 2000; Woltering et al. 2000; Armstrong et al. 2002; Blat et al. 2002; Cai et al. 2003; Higgins et al. 2005; Lam et al. 2005; Kleckner 2006; Sanchez-Moran et al. 2007; Eichinger and Jentsch 2010; Osman et al. 2011; Ferdous et al. 2012).

The seven high confidence candidates for selection exhibit multiple patterns of differentiation within and between the three population groups. Two genes (*ASY3* and *SYN1*) showed differentiation in all contrasts (table 1 and fig. 2) and reduced diversity in Pannonian diploids and tetraploids, but not Carpathian diploids, a pattern consistent with selection for different polymorphisms in Pannonian diploid and tetraploid lineages (supplementary fig. S3, Supplementary Material online). The alleles targeted by selection in each case were distinct: Regions of reduced diversity in Pannonian diploids and tetraploids do not completely overlap (supplementary fig. S3, Supplementary Material online), there is significantly elevated differentiation between Pannonian diploids and tetraploids (table 1 and fig. 2), and high frequency derived (HFD) polymorphisms are distinct in the two lineages (fig. 3). *SMC1* shows significant differentiation between Carpathian diploids and tetraploids, only marginal differentiation between Pannonian diploids and tetraploids (table 1 and supplementary table S2, Supplementary Material online), and reduced genetic diversity in both diploid lineages (supplementary fig. S3, Supplementary Material online). In contrast to *ASY3* and *SYN1*, the same variants in *SMC1* seem to have been selected in both lineages (table 1 and fig. 3). Four genes (*ASY1, PDS5, ZYP1a*, and *ZYP1b*) showed outlier levels of differentiation between diploids and tetraploids, but not among diploids (table 1 and supplementary table S2, Supplementary Material online; fig. 2 and supplementary fig. S2), and reduced nucleotide diversity only within tetraploids (supplementary fig. S3, Supplementary Material online). All four were previously reported as candidate targets of polyploid-specific selective sweeps in a genome-resequencing study that included a much smaller number of individuals (Yant et al. 2013). For *ZYP1*, which has two adjacent functionally redundant tandem duplicates (Higgins et al. 2005), we find elevated differentiation (fig. 2) and reduced diversity (supplementary fig. S3, Supplementary Material online) primarily in *ZYP1b*. Six genes (*AT1G27900, MEI1, MSH2, PRD3, SMC3*, and *SMC6a*) had weak patterns of differentiation that were inconsistent among different tests of differentiation and/or did not show patterns suggestive of selection within a specific region or cytotype (supplementary table S2, Supplementary Material online). Fifty-five genes had few or no outlier windows in any of the population pools and two genes did not have enough coverage to be included in analyses (supplementary table S2, Supplementary Material online). In order to quantify the possible problem of missing rare variants inherent in sequencing pools of genomic DNA, we generated and sequenced long-range polymerase chain

Table 1. Genes with Significant Outlier Values for Three Differentiation Statistics.

				F_{ST}				G				DD				
Gene	ATG ID	Contr.	Bins/ Gene	Out. Bins	*P* Val	Max Bin	Max Rnk	Out. Bins	*P* Val	Max Bin	Max Rnk	Out. Bins	*P* Val	Min Bin	Min Rnk	Type
		C × P	19	3	**	0.46	*	1	—	41.2	—	3	**	−0.06	**	
ASY3	AT2G46980	C × Tet	20	17	****	0.58	***	17	****	61.4	***	16	****	−0.20	***	ALL
		P × Tet	14	10	****	0.59	**	11	****	64.2	**	9	****	−0.10	***	
		C × P	23	11	****	0.65	**	9	****	88.0	***	8	****	−0.09	***	
SYN1	AT5G05490	C × Tet	27	12	****	0.35	***	16	****	32.5	***	9	****	−0.16	**	ALL
		P × Tet	20	9	****	0.81	****	8	****	91.1	***	7	****	−0.15	***	
		C × P	28	1	—	0.53	**	1	—	56.6	**	1	—	−0.10	**	
SMC1	AT3G54670	C × Tet	33	15	****	0.33	***	16	****	46.4	***	15	****	−0.24	***	2×
		P × Tet	25	0	—	0.42	—	6	****	39.9	*	3	**	−0.10	*	
		C × P	25	0	—	0.40	—	2	—	44.3	*	2	—	−0.13	**	
ASY1	At1g67370	C × Tet	26	9	****	0.34	***	10	****	30.0	**	10	****	−0.17	***	4×
		P × Tet	22	3	**	0.53	***	5	****	59.6	**	5	****	−0.15	**	
		C × P	32	0	—	0.38	—	0	—	37.9	—	1	—	−0.10	**	
PDS5	AT1G77600	C × Tet	31	26	****	0.59	***	27	****	67.2	***	25	****	−0.29	***	4×
		P × Tet	24	3	**	0.56	**	8	****	57.4	**	16	****	−0.19	***	
		C × P	16	0	—	0.18	—	0	—	24.1	—	0	—	−0.04	—	
ZYP1a	At1g22260	C × Tet	12	7	****	0.28	**	7	****	37.0	***	6	****	−0.13	**	4×
		P × Tet	10	0	—	0.39	—	0	—	34.2	—	0	—	−0.05	—	
		C × P	27	0	—	0.19	—	0	—	15.9	—	0	—	−0.05	—	
ZYP1b	At1g22275	C × Tet	19	17	****	0.52	***	17	****	39.5	***	14	****	−0.22	***	4×
		P × Tet	14	1	—	0.46	*	2	*	42.6	**	5	****	−0.14	**	

NOTE.—Gene name indicates common name and ATG ID indicates the *Arabidopsis thaliana* gene ID number (www.arabidopsis.org; last accessed 24 December, 2014). Contr., contrast of population groups where C is Carpathian diploid, P is Pannonian diploid, and Tet is tetraploid. Bins are 40 SNP windows and Bins/Gene indicates how many bins lie in each gene region. "Out. Bins" indicates the number of outlier bins (regions which exceed the 1% genome-wide threshold). "*P* Val" indicates whether the number of outlier bins per gene is significantly enriched compared with genome-wide expectation where — > 0.01, *0.01–0.005, **0.005–0.0005, ***0.0005–0.00001, and ****<0.00001. "Max bin" gives the maximum value of the F_{ST}, *G*, and DD statistic for any bin overlapping the focal gene and "Max Rnk" gives genome wide rank for the bin with symbols as for *P* value.

reaction (Ir-PCR) products from a subset of three strongly selected candidate genes (*ASY1*, *ASY3*, and *ZYP1b*) and flanking regions in 113 tetraploids, 21 Carpathian diploids, and 42 Pannonian diploids from 17 populations (supplementary table S1, Supplementary Material online). From this data set, we were able to confirm that similar patterns of population differentiation exist for these genes in both data sets (data not shown). We demonstrate from this also that many (though not all) high frequency polymorphisms characteristic of one lineage are present as rare variants in at least one other lineage (supplementary fig. S4 and table S4, Supplementary Material online). This pattern maybe caused by ancestral segregating variation or polymorphisms that introgressed through more recent gene flow.

High Frequency Amino Acid Changes in Meiosis Genes

We next examined the patterns of HFD amino acid changes relative to the *A. lyrata* reference genome (Hu et al. 2011). The patterns agree with our prior estimations of which lineages experienced selection. In the two genes that show evidence of selection in Pannonian diploids as well as tetraploids (*ASY3* and *SYN1*), we see distinct sets of HFD amino acid changes (>0.5 allele frequency differences among populations) in Pannonian diploids and tetraploids, whereas Carpathian diploids have sequences very similar to *A. lyrata* (fig. 3 and supplementary table S5, Supplementary Material online). For both genes, the Pannonian diploid and tetraploid lineages each have numerous distinct HFD amino acid changes (fig. 3). Of the four *SMC1* HFD amino acids, three (V614I, L706Q, and K911R) are shared between the Carpathian and Pannonian diploid lineages and one (E795D) is restricted to the Carpathian lineage (fig. 3 and supplementary table S5, Supplementary Material online). For *ASY1*, *PDS5*, *ZYP1a*, and *ZYP1b*, the majority of HFD nonsynomous changes are unique to the tetraploid lineage (fig. 3 and supplementary table S5, Supplementary Material online).

We next examined whether HFD amino acid changes in five proteins under selection (*SYN1*, *ASY3*, *ASY1*, *PDS5*, and *SMC1*) affect sites that are conserved in angiosperms and/or whether they are located in known protein functional domains. We excluded *ZYP1a* and *ZYP1b* as the duplication of these genes in the genus makes alignment and SNP conservation calls ambiguous. We analyzed alignments of the closest homologs of these five genes from 31 plant species. Out of 113 HFD differences across the five genes and three population groups in *A. arenosa*, only seven affect strongly conserved (>85% identity across all sampled species) amino acids (supplementary table S6, Supplementary Material online). All of these are tetraploid-specific: One is in the DNA-binding HORMA domain of *ASY1* (K40E), two in *ASY3* (T265I and

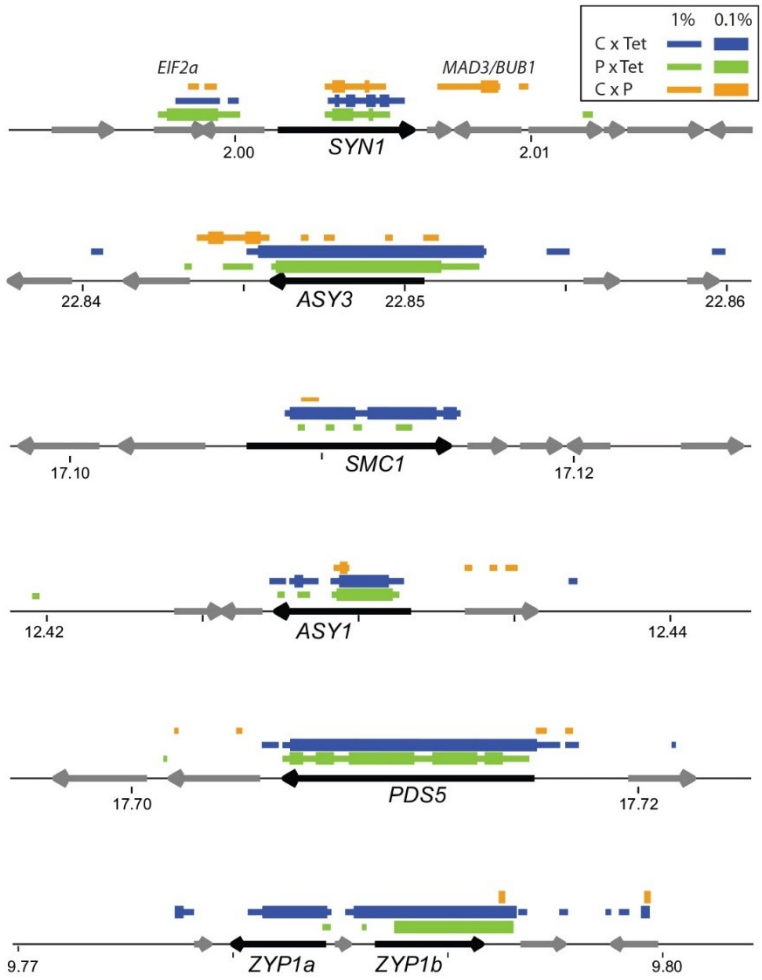


Fig. 2. Plots of DD outlier regions in eight meiosis genes. Plots indicating the spans of 40 SNP windows with outlier ranks (thin lines = 1%, thick lines = 0.1%), with target genes indicated with black arrows and other genes in the region with gray arrows. Blue lines, Carpathian diploid versus tetraploid (C × Tet) contrast; green lines, Pannonian diploid versus tetraploid (P × Tet) contrast; orange lines, Carpathian diploid versus Pannonian diploid (C × P) contrast. DD is calculated for ASY3, SYN1, ZYP1, PDS5 as a function of diversity in Pannonian diploids for C × P, and in tetraploids for C × Tet and P × Tet. DD is calculated for ASY1 as a function of diversity in Carpathian diploids for C × P, and in tetraploids for C × Tet, P × Tet. DD is calculated for SMC1 as a function of diversity in Carpathian diploids in C × P, Carpathian diploids in C × Tet, for Pannonian diploids in P × Tet. Gene models in each region are given as arrows at the bottom of each chart with the meiosis gene of interest in black and labeled.

L268V), two in *PDS5* (S242F and S527Y), and two in *SMC1* (F595S and Q923K) (fig. 3). Several HFD changes alter charged amino acids (D, E, K, and R). Some genes have relatively few charge-altering HFD sites (4/23 in *PDS5*, 4/20 in *SYN1*, 1/10 in *SMC1*, and 9/42 in ASY3), whereas *ASY1* has a larger fraction (3/7). In two cases charge switches completely between negative and positive residues, rather than from charged to uncharged, once in the tetraploid allele of *PDS5* (E949K), and once in the HORMA domain of *ASY1* in tetraploids (K40E).

We also asked whether any HFD amino acid polymorphisms map to known functional domains. The HORMA and SWIRM domains in *ASY1* have one and two HFD amino acid changes each, respectively (fig. 3). *ASY1* and *ASY3* are known to interact with one another, and one of the domains necessary for this interaction has been mapped in *A. thaliana* to the C-terminal tail of *ASY3* (Ferdous et al. 2012). Though *ASY3* is riddled with 42 HFD amino acid differences among population pools, not one of these falls within this *ASY1*-interaction domain (fig. 3). The core of the sister chromatid cohesin complex comprised three structural proteins (*SYN1*, *SMC1*, and *SMC3*), which form a ring enclosing and stabilizing the sister chromatids (Cai et al. 2003; Haering et al. 2004). In the yeast homolog of *SYN1*, *Rec8*, the N and C terminal domains bind *SMC1* and *SMC3* (Haering et al.

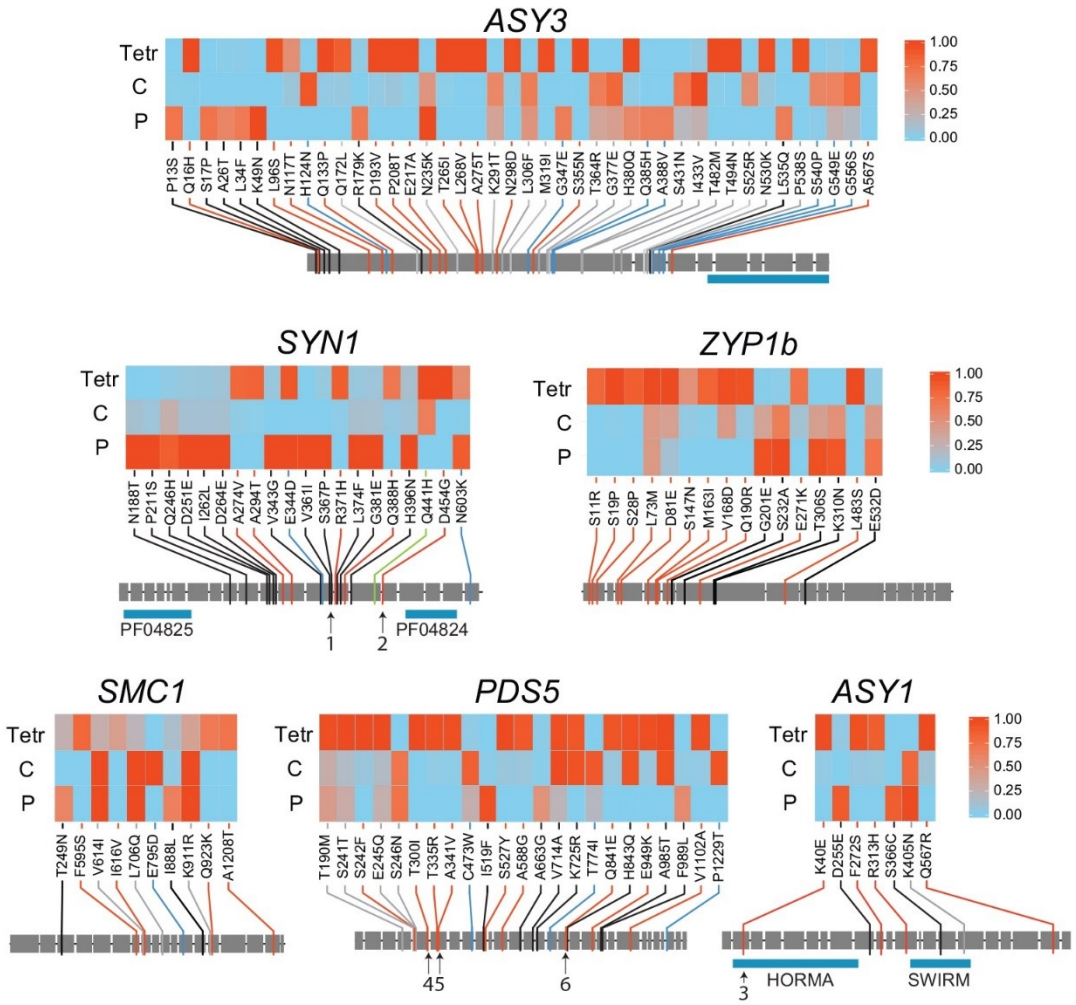


Fig. 3. HFD amino acid polymorphisms in meiosis genes in *Arabidopsis arenosa*. Tetr, tetraploid; C, Carpathian diploid; P, Pannonian diploid. Each gene model is indicated with gray bars, with intron locations indicated by gaps. Functional domains are indicated by blue bars below gene models. The ASY3 region is a portion of the protein known to interact with ASY1. Red lines show sites with unique derived alleles in tetraploids, black with unique derived alleles in Pannonian diploids, green with derived alleles in both Carpathian diploids and tetraploids, and gray indicates other patterns. Numbers indicate highlighted polymorphisms: 1) S367P in SYN1 in Pannonian diploids is predicted to cause loss of serine phosphorylation at S367 and gain of serine phosphorylation at adjacent S366, 2) D454G in SYN1 in tetraploids is predicted to cause loss of phosphorylation at the highly conserved S458, 3) K40E in ASY1 in tetraploids causes a charge switch at a highly conserved amino acid in the HORMA domain adjacent to a Tyrosine predicted by NetPhos to be phosphorylated, 4) Mutation T300I in PDS5 causes loss in tetraploids of a predicted ATM phosphorylation site, 5) two high frequency mutations in PDS5, T335R and A341V, are predicted to cause loss of phosphorylation at Y336 and Y340 in tetraploids, and 6) Q841E in PDS5 in tetraploids is predicted to cause the gain of a sumoylation site centered on residue 839.

2002). These domains are recognizable in the *Arabidopsis* homologs and both domains lack HFD amino acid differences among *A. arenosa* populations.

Finally, as many of the meiosis proteins we found under selection are regulated by phosphorylation and sumoylation (e.g., Spirek et al. 2009; Eichinger and Jentsch 2010; Watts and Hoffmann 2011; Voelkel-Meiman et al. 2013; Zhang et al. 2014), we asked whether any of the HFDs in either Pannonian diploids or tetraploids might affect sites important for this. We mapped sites with four phosphorylation predictors and one sumoylation predictor. In only some cases do predicted phosphorylation sites overlap among different predictors; therefore, we limited ourselves to 43 "high confidence" sites across five proteins that were predicted to be phosphorylated by at least three of the programs (supplementary table S7, Supplementary Material online). High

confidence phosphorylation and sumoylation sites generally do not differ among *A. arenosa* populations, but there are three exceptions, all of which are predicted to cause losses of phosphorylation sites in tetraploids: The nearly fixed tetraploid-specific mutation D454G is predicted to cause loss of phosphorylation at S458 in *SYN1*, and two predicted tyrosine phosphorylation sites in *PDS5* (Y336 and Y340) are predicted to be inactivated by two nearby mutations that are nearly fixed in tetraploids (T335R and A341V; fig. 3). The two *PDS5* tyrosine sites are not well conserved across angiosperms, but S458 in *SYN1* is nearly perfectly conserved (in 26/28 homologous sequences). In no case is a predicted sumoylation site destroyed by HFD amino acid changes in any *A. arenosa* lineage, but the Q841E mutation found in the tetraploid allele of *PDS5* is predicted to introduce a new weak sumoylation site.

Discussion

The core processes of meiosis are conserved across eukaryotes (Villeneuve and Hillers 2001; Gerton and Hawley 2005; Mercier and Grelon 2008; Harrison et al. 2010). Nevertheless, there is growing evidence from studies in animals that there is a surprising degree of diversity in genes that encode structural proteins essential for key meiotic functions, including chromosome pairing, repair, and crossing over (e.g., Williamson et al. 2007; Turner et al. 2008; Anderson et al. 2009; Cheng et al. 2012). As we previously found evidence that tetraploid *A. arenosa* experienced selection on meiosis genes (Hollister et al. 2012; Yant et al. 2013), here we studied the molecular evolution of 70 meiosis genes in both diploid and tetraploid lineages in more detail. We sampled six populations from the single autotetraploid lineage as well as six diploid populations. Among diploids, we find that those from the Carpathian Mountains are much more closely related to the autotetraploids than those from the adjacent, but biogeographically distinct Pannonian Basin.

We found seven meiosis genes with strong signatures suggestive of selection in at least one *A. arenosa* lineage. Three meiosis genes showed evidence that selection acted on them in diploid lineages, and two of these (*ASY3* and *SYN1*) underwent independent selective sweeps targeting different alleles in both Pannonian diploids and in tetraploids. The finding of selection on meiosis genes in diploid *A. arenosa* parallels previous observations made in insects that a subset of meiosis genes involved in chromosome alignment, synapsis, and crossing-over can be strongly differentiated among populations (Turner et al. 2008; Anderson et al. 2009; Cheng et al. 2012). What drives the evolution of meiosis functioning proteins when the basic functions of meiosis are conserved across eukaryotes remains uncertain. However, as we discuss below, there are several hypotheses that may explain these patterns.

One possibility is that divergence is functionally neutral and is favored not because of selection on meiotic processes themselves. In this scenario, divergence is driven by coevolution such that a mutation in one protein that rises in frequency (either due to drift or selection) will cause selection for compensatory mutations in interacting partners to reoptimize protein complexes or interactions. This process can drive rapid divergence of multiple genes encoding members of protein complexes (e.g., Castillo-Davis et al. 2004; Takahashi et al. 2014). There is empirical evidence that such coevolutionary dynamics could be important for meiosis proteins. Several proteins with core functions in DNA repair and homologous recombination show molecular signatures suggestive of coevolution in fungi and animals (e.g., Clark et al. 2013). Furthermore, in natural strains of *Saccharomyces cereviseae*, allelic divergence in two interacting DNA repair proteins, MLH1 and PMS1, yields complexes that are fully functional in their native genomic context, but when brought together in hybrids, the alternate derived alleles cannot properly interact in heteromeric complexes, and very high mutation rates result (Heck et al. 2006; Demogines et al. 2008). In another study, expression of a wild-type RAD52 protein from *Kluyveromyces lactis* in *S. cereviseae* disrupts the interaction of the native RAD52 with its partner RAD51 and impairs DSB repair and recombination (Milne and Weaver 1993). As the proteins we found have high rates of differentiation between *A. arenosa* lineages interact to coordinate a common process, it is possible that coevolution could drive the divergence we observe in interacting proteins between populations, whether or not there is direct selection for alterations in core meiotic processes.

An additional explanation, not mutually exclusive to coevolutionary divergence, is that there may also be direct selection on genome-wide CO rates. Theoretical models suggest that higher recombination rates can promote adaptability of populations by increasing the likelihood that beneficial mutations are fixed and deleterious variants removed (e.g., Otto and Barton 1997; Marais and Charlesworth 2003; Roze and Barton 2006). Related to this is the possibility is that higher CO rates can be favored in order to suppress meiotic drivers (e.g., Haig and Grafen 1991). There is no information on CO frequency in the Pannonian diploids, so we cannot assess whether this hypothesis may explain the divergence in *ASY3* and *SYN1* between diploid lineages.

For the autotetraploid *A. arenosa* lineage, direct selection on CO rates may have occurred in association with the genome duplication in this lineage. Reductions in CO number, ideally to one per chromosome, can contribute to meiotic stability in autopolyploids by preventing the formation of multivalent associations among the multiple homologs, which correlate negatively with fertility (e.g., Hazarika and Rees 1967; Wolf et al. 1989; Khawaja et al. 1997; Cifuentes et al. 2010; Bomblies and Madlung 2014; Lavania). Like many other naturally occurring autotetraploids, *A. arenosa* has reduced CO numbers relative to its diploid progenitor (Comai et al. 2003; Yant et al. 2013). The genes we found here and previously (Hollister et al. 2012; Yant et al. 2013) to have evidence of selection in tetraploids (*ASY1, ASY3, PDS5, ZYP1, SYN1*) encode structural proteins that together coordinate key events in early prophase including sister chromatid cohesion, axis formation, and synapsis (Hollingsworth et al. 1990; Bai et al. 1999; Panizza et al. 2000; Woltering et al. 2000; Armstrong et al. 2002; Blat et al. 2002; Cai et al. 2003; Higgins et al. 2005; Lam et al. 2005; Kleckner 2006; Sanchez-Moran et al. 2007; Eichinger and Jentsch 2010; Osman et al. 2011; Ferdous et al. 2012). Importantly, these proteins are all directly or indirectly involved in coordinating CO formation

(Hollingsworth et al. 1990; Woltering et al. 2000; Armstrong et al. 2002; Higgins et al. 2005; Sanchez-Moran et al. 2007; Osman et al. 2011; Ferdous et al. 2012). Transgenic modulation of expression levels suggests that at least one of these genes (*ASY1*) can affect CO rates and the frequency of multivalent formation in tetraploid wheat (Boden et al. 2009). We thus hypothesize that selection on meiosis genes in the tetraploid *A. arenosa* lineage may reflect the need to reduce CO rates to prevent multivalents and promote fertility in a tetraploid genomic context. Whether the observation that more genes and more amino acids at conserved sites are differentiated in the tetraploid, relative to the diploid lineage translates to a greater effect on meiosis in this lineage will be interesting to test.

For the Pannonian diploid lineage, there is an additional possibility, also not mutually exclusive with aforementioned hypotheses. The Pannonian diploids are from a distinct biogeographic region (Molnár et al. 2006; Erdös et al. 2014) that Worldclim data show is significantly warmer than surrounding mountain habitats of the tetraploids and the Carpathian diploids. Meiosis is known to be acutely sensitive to temperature in natural populations of both plants and animals, and this is particularly true of the very processes controlled by the genes that are strongly differentiated between *A. arenosa* lineages (axis formation, synapsis and CO regulation) (Wilson 1959a; Nebel and Hackett 1961; Buss and Henderson 1988; Henderson 1988; Loidl 1989; Bilgir et al. 2013). Elevated temperatures, sometimes of just a few degrees above normal, can lead to failures of synapsis and associated CO formation (Wilson 1959b, 1959a; Nebel and Hackett 1961; Shaw 1974; Buss and Henderson 1988; Henderson 1988; Loidl 1989; Bilgir et al. 2013). Importantly, the threshold temperatures at which meiosis becomes unstable, and the degree of sensitivity to temperature, are known to vary among wild populations and this variation has a genetic basis in at least some cases (e.g., Wilson 1959b; Riley et al. 1966; Shaw 1974; Buss and Henderson 1988; Francis et al. 2007). Taken together, these findings raise the interesting possibility that as populations colonize novel habitats or as global climate change alters the temperature in native habitats, there may be selection to stabilize the physical structures and protein interactions critical for meiotic axis formation, synapsis, and CO regulation. Whether temperature can be a driver of divergence in meiosis genes remains to be tested.

Whether the underlying causes of selection on meiosis genes in the two *A. arenosa* lineages we studied are different remains to be seen. If both diploid and tetraploid lineages are under selection for example for reduced CO rates, then it would be unsurprising that similar sets of genes are targeted. However, if the underlying causes are different, the observation that overlapping functional classes, and in two cases, distinct alleles of the same genes, are strongly differentiated among lineages could suggest that there are relatively few ways in which meiotic stabilization can be achieved by natural selection when the system is faced with challenges. As similarly sampled genome scans with molecular follow-up studies become available in this and other species, it will be interesting to ask whether the same sets of genes or functional classes are repeatedly under selection in tetraploid and/or diploid lineages and what the causal selective factors are.

Materials and Methods

Plant Material

We grew plants from first-generation seeds collected from wild populations in June/July of 2010 and 2012 (supplementary table S1, Supplementary Material online). We grew plants in growth chambers (Conviron MTPC144) under 18-h days, with 12 °C night temperatures and 18–22 °C day temperatures. We extracted DNA from individuals using a modified CTAB protocol as previously described (Hollister et al. 2012; Yant et al. 2013).

Sequencing and Variant Identification

PoolSeq Data set

We conducted pooled whole-genome resequencing from 12 populations, including 3 Pannonian diploid populations, 3 Carpathian diploid populations, and 6 tetraploid populations. We sampled 8–26 individuals from each population (supplementary table S1, Supplementary Material online) and pooled equal quantities of DNA for each individual in a population. For each pool, we prepared a barcoded DNA library for sequencing using adapter and primer combinations modified from (Peterson et al. 2012). We sequenced all ten libraries to a minimum of 11× coverage on three Illumina HiSeq lanes (supplementary table S8, Supplementary Material online). Three very closely related railway tetraploid populations (STE, TBG, and NOW) had lower than expected coverage (1.3–8.3×) compared with the other populations and we thus pooled these into a single population. We aligned reads to the *A. lyrata* genome (Hu et al. 2011) using the program STAMPY (Lunter and Goodson 2011). We used raw read number to call allele frequency with SNAPE (Raineri et al. 2012) to calculate a maximum-likelihood estimate of allele frequency, accounting for variation in read quality and coverage. For a site to be included in our analyses of population differentiation, we required a minimum of 15× coverage per population and a maximum coverage of 100×. We next created an additional data filtration step to exclude loci at which reads mapped from potentially paralogous loci as follows: Using eight diploid whole-genome sequences (Yant et al. 2013), we identified loci with erroneously mapped reads at annotated genes or intergenic segments (no longer than 2 kb) in which all eight diploids were heterozygous at more than two sites within the defined region. We excluded these loci from all downstream analyses, as it is unlikely for all eight diploids from two distinct populations to be heterozygous at three or more sites within a gene or intergenic segment according to Hardy–Weinberg equilibrium. We used only diploid genomes for this analysis in order to prevent ascertainment bias against potentially diploidized loci in the autotetraploids. We use this filtered data set in our pairwise population analyses; the number of sites and SNPs for each analysis is provided in supplementary table S9, Supplementary Material online. For a site to be included in our analysis of population group differentiation, coverage

must exceed 25×. After this filtering we identified a total of 9.45 million SNPs for the Carpathian and Pannonian diploid contrasts (C × P), 8.71 million SNPs for the Pannonian diploid by tetraploid contrast (P × Tet), and 11.74 million SNPs for the Carpathian diploid by tetraploid contrast (C × Tet).

Ir-PCR Data Set

To sequence *ASY1*, *ASY3*, and *ZYP1* from a larger set of individuals, we amplified the genes by Ir-PCR using Phusion High Fidelity DNA Polymerase (New England Biolabs) from 42 Pannonian diploids, 21 Carpathian diploids, and 113 tetraploids. We used primers spanning the genes and adjacent intergenic regions conserved between *A. arenosa* and *A. thaliana* to generate products of 7.1 kb for *ASY1*, 17 kb for *ASY3*, and 9.5 kb for *ZYP1b*. We pooled PCR product, barcoded each sample as in Peterson et al. (2012), and then sequenced all products together. We aligned reads to the *A. lyrata* genome (Hu et al. 2011) using STAMPY (Lunter and Goodson 2011), and called polymorphisms using GATK (McKenna et al. 2010), following GATK "Best Practices" protocols.

Population Relatedness and Differentiation

We conducted a PCA on variant sites that exceed the minimum 15× coverage in at least seven of ten populations (290,000 sites), randomly subsampled this data set to 50,000 sites 100 times, and conducted PCA on each subsampled data set. As these data sets contain missing data, we conducted the PCA using a probabilistic model to estimate missing values in the "ppca" program in the R package pcaMethods (Stacklies et al. 2007).

We assessed the degree of differentiation between populations using three different statistics: F_{ST} (Weir and Cockerham 1984), *G* (Magwene et al. 2011), and DD (Yant et al. 2013). We calculate F_{ST} as $\pi_B - \pi_W/\pi_B$ (Weir and Cockerham 1984); π_W is the within-population expected heterozygosity, $2p(1-p)$, where p is the frequency of the reference allele within a population. π_B is the between-population heterozygosity and is calculated using the same formula except that here p is the weighted average allele frequency in both populations. In order to compare heterozygosity between cytotypes, we calculate the "gametic heterozygosity" of tetraploid populations (Moody et al. 1993). Gametic heterozygosity of an autotetraploid population and the homozygosity of a diploid population each define the probability that a pair of homologous sites, randomly drawn without replacement from a randomly selected individual, is identical (Moody et al. 1993). We also calculate a related measure of population differentiation that accounts for the amount of genetic diversity in a particular genomic window, the DD statistic, as the residual deviation from a linear regression fit to the relationship between differentiation ($p_i - p_j$) and diversity in one population (π_{Wi}) where p is the frequency of the reference allele in population i or j and π_{Wi} is the expected heterozygosity in population i (Yant et al. 2013). For each pairwise population contrast, we also calculate the reciprocal DD statistic, DD′ as taking the residual value from a linear regression fit to $p_i - p_j$ and π_{Wj}. A large difference between DD and DD′, a result of low heterozygosity in one population relative to the other, suggests a selective sweep may have occurred in one population. Because our goal is to identify candidate selective sweeps, we report the larger of these two values for a particular genomic interval. Finally, we calculate the *G* statistic as a measure of population differentiation weighted by coverage at each site in each population. The *G* statistic is: $2 \times \Sigma n_i \ln(n_i/n_{exp})$ where, n_i is observed allele count, i is each allele state at a particular site, and n_{exp} is the expected allele count assuming no difference in allele frequency between populations (Magwene et al. 2011). We excluded all sites where n_{exp} is less than 2.

We calculated each measure of differentiation between all possible pairs of populations ($n = 45$) and for the three population pools identified in the PCA analysis: Carpathian diploids, Pannonian diploids, and tetraploids. For pooled population data sets, we increased our filtering stringency so coverage at a site had to exceed 20× and the minor allele frequency of a putative SNP across populations had to exceed 0.01. We measured each statistic in 20, 40, and 100 SNP bins and assessed the significance of each window by its location in the genome-wide rank distribution for each independent population comparison.

For each of the 70 meiosis functioning proteins, we measured the total number of SNP bins that reside within ±2 kb of gene start/end sites. We next measured the number of gene-specific bins which reside in 1% tail of the genome-wide distribution of F_{ST}, G, and DD, and determined whether the number of outlier bins in a single gene is enriched compared the genome wide background using a one-sided hypergeometric distribution (Rivals et al. 2007).

Determining Protein Orthologs

We collected amino acid sequences for the meiosis genes we analyzed for conservation from 31 flowering plant species from phytozome (www.phytozome.com; last accessed 24 December, 2014). We included *Oryza sativa, Brachypodium distachyon, Panicum virgatum, Setaria italica, Sorghum bicolor, Zea mays, Aquilegia coerulea, Vitis vinifera, Populus trichocarpa, Linum usitatissimum, Manihot esculenta, Ricinus communis, Medicago truncatula, Glycine max, Prunus persica, Malus domestica, Fragaria vesca, Phaseolus vulgaris, Eucalyptus grandis, Citrus clementina, Citrus sinensis, Theobroma cacao, Gossypium raimondii, A. lyrata, A. thaliana, Capsella rubella, Thellungiella halophila, Brassica rapa, Mimulus guttatus, Solanum lycopersicum*, and *Solanum tuberosum*. We used InParanoid to search for orthologs between species using the BLOSSUM62 matrix with similarity cutoff score of 40 bits, a confidence score greater than 0.05, a sequence overlap cutoff of 0.5, and a group merging cutoff 0.5. The sequences of the 70 meiotic proteins (Yant et al. 2013) from *A. thaliana* were searched against the proteomes from each species to find orthologs, if present. Commonly, multiple orthologs were found to match the query sequence, in which case only the reciprocal best hits were taken as the ortholog matches to simplify further analysis. The sets of orthologous protein sequences were aligned using MUSCLE (Edgar 2004). We visualized and annotated the conserved domain information in

Genious (v7.1) (Kearse et al. 2012) with domain assignments collected from the *Arabidopsis* genome annotation (TAIR; www.arabidopsis.org; last accessed 24 December, 2014).

Phosphorylation and Sumoylation Scans

To identify candidate sumoylation sites, we used GPS-Sumo (http://sumosp.biocuckoo.org; last accessed 24 December, 2014). To identify candidate phosphorylation sites, we used four different servers and limited ourselves to sites that were identified in at least three of them. We used: NetPhos 2.0 (Blom et al. 1999), DiPhos (http://www.dabi.temple.edu/disphos/information.html#links; last accessed 24 December, 2014), Scansite3 (Obenauer et al. 2003), and P3DB (Zanzoni et al. 2007). We used the *A. lyrata* reference sequences for each gene to identify candidate sites, and then altered this sequence with known high frequency changes from either Pannonian diploids or tetraploids to ask whether these changes affect the predicted modifications at identified sites.

Bioclimatic Differentiation

We used the Worldclim data set to find bioclimatic data for all diploid and tetraploid populations (Hijmans et al. 2005). We extracted bioclimatic data at the finest resolution of the Worldclim data set, 30 s × 30 s (~1 km^2) cells using the R package "raster" (Hijmans et al. 2005). We conducted a PCA of all 19 bioclimatic variables in the data set as well as a subset of five temperature associated bioclimatic variables: Temperature for the warmest month, warmest quarter (three month period), driest quarter, annual temperature, and seasonality (standard deviation in weekly measurements of temperature compared with mean annual temperature). We present temperature-specific habitat differentiation (supplementary table S1 and fig. S1, Supplementary Material online), although we find similar levels of differentiation between Carpathian and Pannonian diploid populations in both data sets (results not shown). We tested whether there is a significant effect of population group on each bioclimatic variable using the GLM procedure in R (v. 2.15.1).

Supplementary Material

Supplementary tables S1–S9 and figures S1–S3 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors thank Nancy Kleckner, Dan Hartl, and members of the Hartl and Bomblies labs for helpful advice and discussions, and Levi Yant for helpful comments on the manuscript. They also thank Dina Benayad-Cherif for help with sequencing. This work was supported by an NRSA postdoctoral fellowship from the National Institute of Health (NIH 5F32GM105293) to K.W., and a Harvard University Milton Fund Award and a MacArthur Fellowship to K.B.

References

Anderson JA, Gilliland WD, Langley CH. 2009. Molecular population genetics and evolution of *Drosophila* meiosis genes. *Genetics* 181: 177–185.

Armstrong SJ, Caryl AP, Jones GH, Franklin FCH. 2002. Asy1, a protein required for meiotic chromosome synapsis, localizes to axis-associated chromatin in *Arabidopsis* and *Brassica*. *J Cell Sci.* 115: 3645–3655.

Bai X, Peirson BN, Dong F, Xue C, Makaroff CA. 1999. Isolation and characterization of SYN1, a RAD21-like gene essential for meiosis in *Arabidopsis*. *Plant Cell* 11:417–430.

Bilgir C, Dombecki CR, Chen PF, Villeneuve AM, Nabeshima K. 2013. Assembly of the synaptonemal complex is a highly temperature-sensitive process that is supported by PGL-1 during *Caenorhabditis elegans* meiosis. *G3* 3:585–595.

Blat Y, Protacio RU, Hunter N, Kleckner N. 2002. Physical and functional interactions among basic chromosome organizational features govern early steps of meiotic chiasma formation. *Cell* 111: 791–802.

Blom N, Gammeltoft S, Brunak S. 1999. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol.* 294:1351–1362.

Boden SA, Langridge P, Spangenberg G, Able JA. 2009. TaASY1 promotes homologous chromosome interactions and is affected by deletion of Ph1. *Plant J.* 57:487–497.

Boettger LM, Handsaker RE, Zody MC, McCarroll SA. 2012. Structural haplotypes and recent evolution of the human 17q21.31 region. *Nat Genet.* 44:881–885.

Bomblies K, Madlung A. 2014. Polyploidy in the *Arabidopsis* genus. *Chromosome Res.* 22:117–134.

Bremer G, Bremer-Reinders DE. 1954. Breeding of tetraploid rye in the Netherlands. I. Methods and cytological observations. *Euphytica* 3: 49–63.

Buss ME, Henderson SA. 1988. The effects of elevated temperature on chiasma formation in *Locusta migratoria*. *Chromosoma* 97:235–246.

Cai X, Dong F, Edelmann RE, Makaroff CA. 2003. The *Arabidopsis* SYN1 cohesin protein is required for sister chromatid arm cohesion and homologous chromosome pairing. *J Cell Sci.* 116:2999–3007.

Castillo-Davis CI, Kondrashov FA, Hartl DL, Kulathinal RJ. 2004. The functional genomic distribution of protein divergence in two animal phyla: coevolution, genomic conflict, and constraint. *Genome Res.* 14:802–811.

Cheng C, White BJ, Kamdem C, Mockaitis K, Costantini C, Hahn MW, Besansky NJ. 2012. Ecological genomics of *Anopheles gambiae* along a latitudinal cline: a population-resequencing approach. *Genetics* 190:1417–1432.

Chowdhury R, Bois PRJ, Feingold E, Sherman SL, Cheung VG. 2009. Genetic analysis of variation in human meiotic recombination. *PLoS Genet.* 5:e1000648.

Cifuentes M, Grandont L, Moore G, Chèvre AM, Jenczewski E. 2010. Genetic regulation of meiosis in polyploid species: new insights into an old question. *New Phytol.* 186:29–36.

Clark NL, Alani E, Aquadro CF. 2013. Evolutionary rate covariation in meiotic proteins results from fluctuating evolutionary pressure in yeasts and mammals. *Genetics* 193:529–538.

Comai L. 2005. The advantages and disadvantages of being polyploid. *Nat Rev Genet.* 6:836–846.

Comai L, Tyagi AP, Lysak MA. 2003. FISH analysis of meiosis in *Arabidopsis* allopolyploids. *Chromosome Res.* 11:217–226.

Demogines A, Wong A, Aquadro C, Alani E. 2008. Incompatibilities involving yeast mismatch repair genes: a role for genetic modifiers and implications for disease penetrance and variation in genomic mutation rates. *PLoS Genet.* 4:e1000103.

Doyle JJ, Flagel LE, Paterson AH, Rapp RA, Soltis DE, Soltis PS, Wendel JF. 2008. Evolutionary genetics of genome merger and doubling in plants. *Annu Rev Genet.* 42:443–461.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792–1797.

Eichinger CS, Jentsch S. 2010. Synaptonemal complex formation and meiotic checkpoint signaling are linked to the lateral element protein Red1. *Proc Natl Acad Sci U S A.* 107:11370–11375.

Erdős L, Tölgyesi C, Horzse M, Tolnay D, Hurton.Á, Schulcz N, Körmöczi L, Lengyel A, Bátori Z. 2014. Habitat complexity of the Pannonian

forest-steppe zone and its nature conservation implications. *Ecol Complex.* 17:107–118.
Ferdous M, Higgins JD, Osman K, Lambing C, Roitinger E, Mechtler K, Armstrong SJ, Perry R, Pradillo M, Cuñado M, et al. 2012. Interhomolog crossing-over and synapsis in *Arabidopsis* meiosis are dependent on the chromosome axis protein AtASY3. *PLoS Genet.* 8:e1002507.
Fledel-Alon A, Leffler EM, Guan Y, Stephens M, Coop G, Przeworski M. 2011. Variation in human recombination rates and its genetic determinants. *PLoS One* 6:e20321.
Francis KE, Lam SY, Harrison BD, Bey AL, Berchowitz LE, Copenhaver GP. 2007. Pollen tetrad-based visual assay for meiotic recombination in *Arabidopsis*. *Proc Natl Acad Sci U S A.* 104:3913–3918.
Gerton JL, Hawley RS. 2005. Homologous chromosome interactions in meiosis: diversity amidst conservation. *Nat Rev Genet.* 6:477–487.
Gilles A, Randolph LF. 1951. Reduction of quadrivalent frequency in autotetraploid maize during a period of ten years. *Am J Bot.* 38: 12–17.
Haering CH, Löwe J, Hochwagen A, Nasmyth K. 2002. Molecular architecture of SMC proteins and the yeast cohesin complex. *Mol Cell.* 9: 773–788.
Haering CH, Schoffnegger D, Nishino T, Helmhart W, Nasmyth K, Löwe J. 2004. Structure and stability of cohesin's Smc1-kleisin interaction. *Mol Cell.* 15:951–964.
Haig D, Grafen A. 1991. Genetic scrambling as a defence against meiotic drive. *J Theor Biol.* 153:531–558.
Harrison CJ, Alvey E, Henderson IR. 2010. Meiosis in flowering plants and other green organisms. *J Exp Bot.* 61:2863–2875.
Hazarika MH, Rees H. 1967. Genotypic control of chromosome behaviour in rye. X. Chromosome pairing and fertility in autotetraploids. *Heredity* 22:317–322.
Heck JA, Argueso JL, Gemici Z, Reeves RG, Bernard A, Aquadro CF, Alani E. 2006. Negative epistasis between natural variants of the *Saccharomyces cerevisiae* MLH1 and PMS1 genes results in a defect in mismatch repair. *Proc Natl Acad Sci U S A.* 103:3256–3261.
Henderson SA. 1988. Four effects of elevated temperature on chiasma formation in the locust *Schistocerca gregaria*. *Heredity* 60: 387–401.
Higgins JD, Sanchez-Moran E, Armstrong SJ, Jones GH, Franklin FCH. 2005. The *Arabidopsis* synaptonemal complex protein ZYP1 is required for chromosome synapsis and normal fidelity of crossing over. *Genes Dev.* 19:2488–2500.
Hijmans RJ, Camerson SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol.* 25:1965–1978.
Hollingsworth NM, Goetsch L, Byers B. 1990. The HOP1 gene encodes a meiosis-specific component of yeast chromosomes. *Cell* 61:73–84.
Hollister JD, Arnold BJ, Svedin E, Xue KS, Dilkes BP, Bomblies K. 2012. Genetic adaptation associated with genome-doubling in autotetraploid *Arabidopsis arenosa*. *PLoS Genet.* 8:e1003093.
Hu TT, Pattyn P, Bakker EG, Cao J, Cheng J-F, Clark RM, Fahlgren N, Fawcett JA, Grimwood J, Gundlach H, et al. 2011. The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nat Genet.* 43:476–481.
Jørgensen MH, Ehrich D, Schmickl R, Koch MA, Brysting AK. 2011. Interspecific and interploidal gene flow in Central European *Arabidopsis* (Brassicaceae). *BMC Evol Biol.* 11:346.
Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28: 1647–1649.
Khawaja H, Sybenga J, Ellis JR. 1997. Chromosome pairing and chiasma formation in autopolyploids of different *Lathyrus* species. *Genome* 40:937–944.
Kleckner N. 2006. Chiasma formation: chromatin/axis interplay and the role(s) of the synaptonemal complex. *Chromosoma* 115:175–194.
Kong A, Thorleifsson G, Frigge ML, Masson G, Gudbjartsson DF, Villemoes R, Magnusdottir E, Olafsdottir SB, Thorsteinsdottir U, Stefansson K, et al. 2014. Common and low-frequency variants associated with genome-wide recombination rate. *Nat Genet.* 46: 11–16.
Kong A, Thorleifsson G, Stefansson H, Masson G, Helgason A, Gudbjartsson DF, Jonsdottir GM, Gudjonsson SA, Sverrison S, Thorlacius T, et al. 2008. Sequence variants in the RNF212 gene associate with genome-wide recombination rate. *Science* 319: 1398–1401.
Lam WS, Yang X, Makaroff CA. 2005. Characterization of *Arabidopsis thaliana* SMC1 and SMC3: evidence that AtSMC3 may function beyond chromosome cohesion. *J Cell Sci.* 118:3037–3048.
Lavania UC. 1986. High bivalent frequencies in artificial autopolyploids of *Hyoscyamus muticus* L. *Can J Genet Cytol.* 28:7–11.
Lind-Halldén C, Halldén C, Säll T. 2002. Genetic variation in *Arabidopsis suecica* and its parental species *A. arenosa* and A. thaliana. *Hereditas* 136:45–50.
Loidl J. 1989. Effects of elevated temperature on meiotic chromosome synapsis in *Allium ursinum*. *Chromosoma* 97:449–458.
Lunter G, Goodson M. 2011. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res.* 21: 936–939.
Magwene PM, Willis JH, Kelly JK. 2011. The statistics of bulk segregant analysis using next generation sequencing. *PLoS Comput Biol.* 7:e1002255.
Marais G, Charlesworth B. 2003. Genome evolution: recombination speeds up adaptive evolution. *Curr Biol.* 13:R68–R70.
McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20:1297–1303.
Mercier R, Grelon M. 2008. Meiosis in plants: ten years of gene discovery. *Cytogenet Genome Res.* 120:281–290.
Milne GT, Weaver DT. 1993. Dominant negative alleles of RAD52 reveal a DNA repair/recombination complex including Rad51 and Rad52. *Genes Dev.* 7:1755–1765.
Molnár G, Timár CS, Ferencz C, Székely B, Lichtenberger J, Sallai K. 2006. Land surface temperature (LST) estimations in the Pannonian Basin using MODIS satellite data. *Geophys Res Abstr.* 8:06588.
Moody ME, Mueller LD, Soltis DE. 1993. Genetic variation and random drift in autotetraploid populations. *Genetics* 134:649–657.
Myers WM. 1945. Meiosis in autotetraploid *Lolium perenne* in relation to chromosomal behavior in autopolyploids. *Bot Gaz.* 106: 304–316.
Nebel BR, Hackett EM. 1961. Synaptinemal complexes (cores) in primary spermatocytes of mouse under elevated temperature. *Nature* 190: 467–468.
Obenauer JC, Cantley LC, Yaffe MB. 2003. Scansite 2.0: proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Res.* 31:3635–3641.
Osman K, Higgins JD, Sanchez-Moran E, Armstrong SJ, Franklin FCH. 2011. Pathways to meiotic recombination in *Arabidopsis thaliana*. *New Phytol.* 190:523–544.
Otto SP, Barton NH. 1997. The evolution of recombination: removing the limits to natural selection. *Genetics* 147:879–906.
Otto SP, Whitton J. 2000. Polyploid incidence and evolution. *Annu Rev Genet.* 34:401–437.
Panizza S, Tanaka T, Hochwagen A, Eisenhaber F, Nasmyth K. 2000. Pds5 cooperates with cohesin in maintaining sister chromatid cohesion. *Curr Biol.* 10:1557–1564.
Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One* 7:e37135.
Raineri E, Ferretti L, Esteve-Codina A, Nevado B, Heath S, Pérez-Enciso M. 2012. SNP calling by sequencing pooled samples. *BMC Bioinformatics* 13:239.
Ramsey J, Schemske D. 1998. Neopolyploidy in flowering plants. *Annu Rev Ecol Syst.* 33:589–639.
Rieseberg L, Willis J. 2007. Plant speciation. *Science* 317:910–914.

Riley R, Darlington CD, Lewis KR. 1966. Genotype-environmental interaction affecting chiasma frequency in *Triticum aestivum*. *Chromosomes Today* 1966:57–65.

Rivals I, Personnaz L, Taing L, Potier M-C. 2007. Enrichment or depletion of a GO category within a class of genes: which test? *Bioinformatics* 23:401–407.

Roze D, Barton NH. 2006. The Hill-Robertson effect and the evolution of recombination. *Genetics* 173:1793–1811.

Sanchez-Moran E, Santos JL, Jones GH, Franklin FCH. 2007. ASY1 mediates AtDMC1-dependent interhomolog recombination during meiosis in *Arabidopsis*. *Genes Dev.* 21:2220–2233.

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.

Shaw DD. 1974. Genetic and environmental components of chiasma control. 3. Genetic analysis of chiasma frequency variation in two selected lines of *Schistocerca gregaria* Forsk. *Chromosoma* 46: 365–374.

Soltis DE, Soltis PS, Tate JA. 2003. Advances in the study of polyploidy since *Plant speciation*. *New Phytol.* 161:173–191.

Spirek M, Estreicher A, Csaszar E, Wells J, McFarlane RJ, Watts FZ, Loidl J. 2009. SUMOylation is required for normal development of linear elements and wild-type meiotic recombination in *Schizosaccharomyces pombe*. *Chromosoma* 119:59–72.

Stacklies W, Redestig H, Scholz M, Walther D, Selbig J. 2007. pcaMethods: a bioconductor package providing PCA methods for incomplete data. *Bioinformatics* 23:1164–1167.

Stefansson H, Helgason A, Thorleifsson G, Steinthorsdottir V, Masson G, Barnard J, Baker A, Jonasdottir A, Ingason A, Gudnadottir VG, et al. 2005. A common inversion under selection in Europeans. *Nat Genet.* 37:129–137.

Steinberg KM, Antonacci F, Sudmant PH, Kidd JM, Campbell CD, Vives L, Malig M, Scheinfeldt L, Beggs W, Ibrahim M, et al. 2012. Structural diversity and African origin of the 17q21.31 inversion polymorphism. *Nat Genet.* 44:872–880.

Takahashi Y, Nagata N, Kawata M. 2014. Antagonistic selection factors induce a continuous population divergence in a polymorphism. *Heredity* 112:391–398.

Turner TL, Levine MT, Eckert ML, Begun DJ. 2008. Genomic analysis of adaptive differentiation in *Drosophila melanogaster*. *Genetics* 179: 455–473.

Villeneuve AM, Hillers KJ. 2001. Whence meiosis? *Cell* 106:647–650.

Voelkel-Meiman K, Taylor LF, Mukherjee P, Humphryes N, Tsubouchi H, MacQueen AJ. 2013. SUMO localizes to the central element of synaptonemal complex and is required for the full synapsis of meiotic chromosomes in budding yeast. *PLoS Genet.* 9:e1003837.

Watts FZ, Hoffmann E. 2011. SUMO meets meiosis: an encounter at the synaptonemal complex. *Bioessays* 33:529–537.

Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358.

Williamson SH, Hubisz MJ, Clark AG, Payseur BA, Bustamante CD, Nielsen R. 2007. Localizing recent adaptive evolution in the human genome. *PLoS Genet.* 3:e90.

Wilson JY. 1959a. Temperature effect on chiasma frequency in the bluebell endymion nonscriptus. *Chromosoma* 10:337–354.

Wilson JY. 1959b. Chiasma frequency in relation to temperature. *Genetica* 29:290–303.

Wolf PG, Soltis PS, Soltis DE. 1989. Tetrasomic inheritance and chromosome pairing behaviour in the naturally occurring autotetraploid *Heuchera grossulariifolia* (Saxifragaceae). *Genome* 32: 655–659.

Woltering D, Baumgartner B, Bagchi S, Larkin B, Loidl J, de los Santos T, Hollingsworth NM. 2000. Meiotic segregation, synapsis, and recombination checkpoint functions require physical interaction between the chromosomal proteins Red1p and Hop1p. *Mol Cell Biol.* 20: 6646–6658.

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

Zanzoni A, Ausiello G, Via A, Gherardini PF, Helmer-Citterich M. 2007. Phospho3D: a database of three-dimensional structures of protein phosphorylation sites. *Nucl Acids Res.* 35:D229–D231.

Zhang L, Wang S, Yin S, Hong S, Kim KP, Kleckner N. 2014. Topoisomerase II mediates meiotic crossover interference. *Nature* 511:551–556.

Study 2: Unmasking cryptic biodiversity in polyploids: origin and diversification of *Aster amellus* aggregate

Annals of Botany **122**: 1047–1059, 2018
doi: 10.1093/aob/mcy149, available online at www.academic.oup.com/aob

ANNALS OF BOTANY
Founded 1887

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

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

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

Received: 7 March 2018 Returned for revision: 16 April 2018 Editorial decision: 15 July 2018 Accepted: 18 July 2018
Published electronically 10 August 2018

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

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

INTRODUCTION

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

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

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

© The Author(s) 2018. Published by Oxford University Press on behalf of the Annals of Botany Company.
All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.

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

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

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

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

MATERIALS AND METHODS

Cytotype distribution and evidence for autopolyploidy

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

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

Taxon sampling and DNA sequencing

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

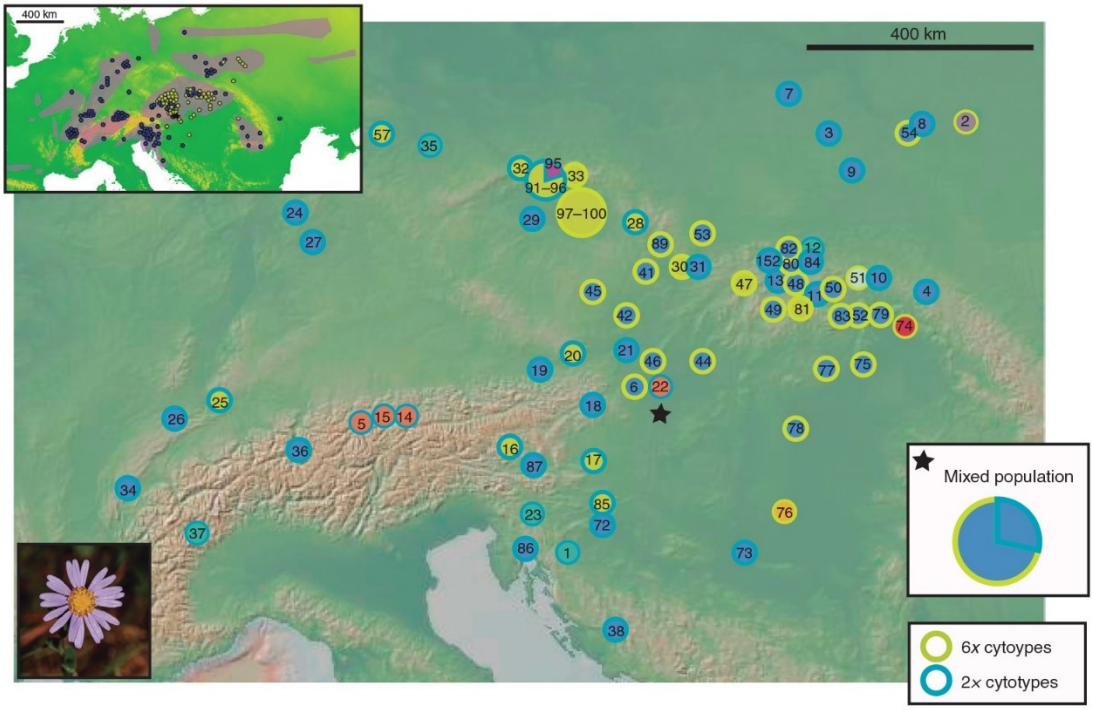


FIG. 1. Geographical location of the 72 populations used in the phylogenetic analyses. Numbers correspond to the sampled populations, with codes given in Supplementary Data Table S1. Pie charts are surrounded in blue for the diploids and in green for the hexaploids, while its circle size is proportional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

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

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

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

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

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

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

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

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

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

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

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

Phylogenetic inference and divergence time estimation

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

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

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

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

Haplotype analyses and demographic history

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

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

Ecological niche modelling

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

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

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

RESULTS

Phylogenetic relationships and molecular dating

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

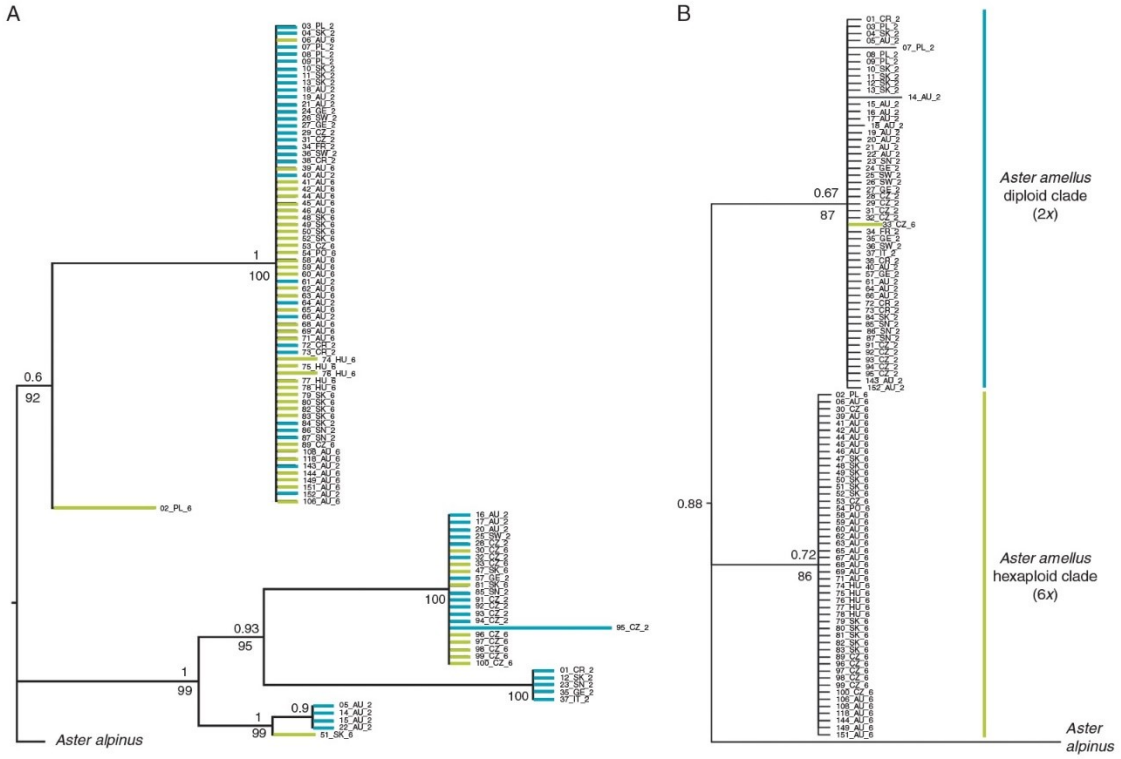


FIG. 2. Bayesian majority-rule consensus trees obtained by MrBayes from: (A) the concatenated *A. amellus* chloroplast data set (*atp*I–*atp*H, *rps*16–*trn*K, *rpl*32–*trn*L, *psb*E–*pet*L and *pet*N–*psb*M); (B) the *A. amellus* nuclear ribosomal (ITS) data set. Numbers above branches indicate Bayesian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

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

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

Haplotype network distribution and demographic analyses

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

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

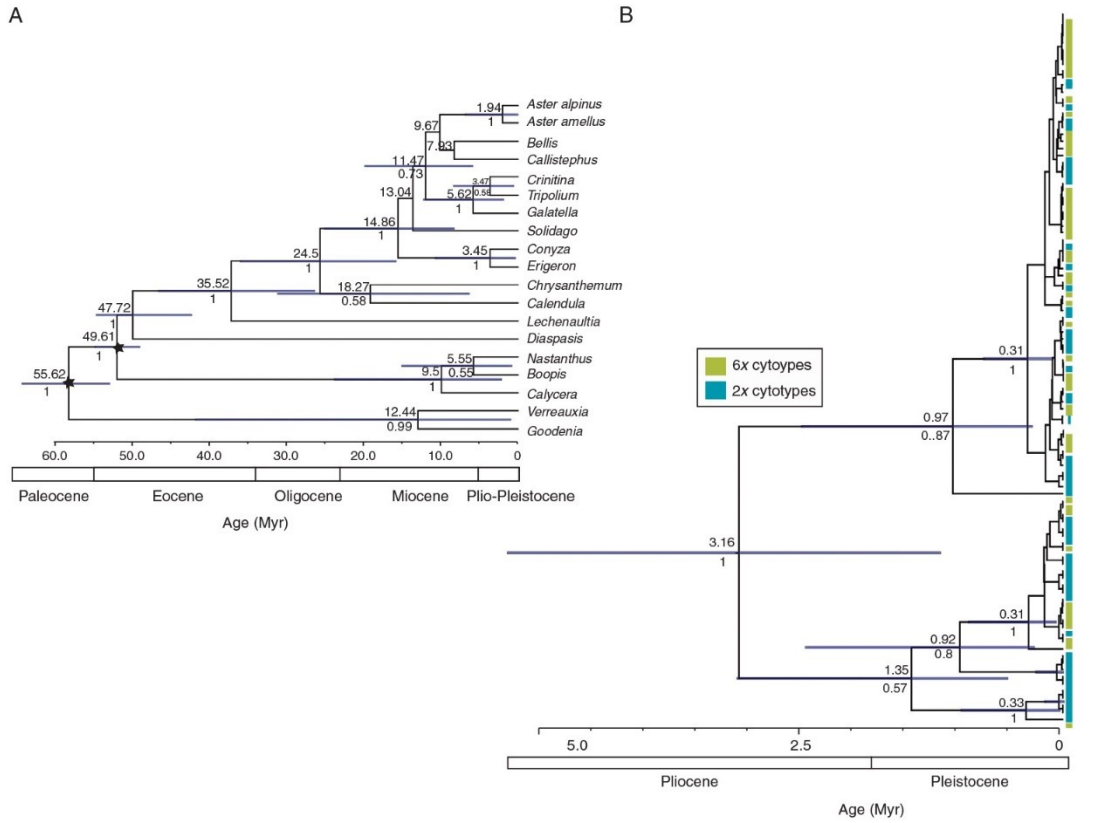


FIG. 3. Bayesian majority-rule consensus trees obtained by the nested analyses of the linked data sets: (A) 'pDNA outgroups data set' and (B) 'amellus pDNA data set'. Stars indicate constrained nodes (see text for more details). Horizontal bars show the 95 % HPD confidence intervals for the supported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

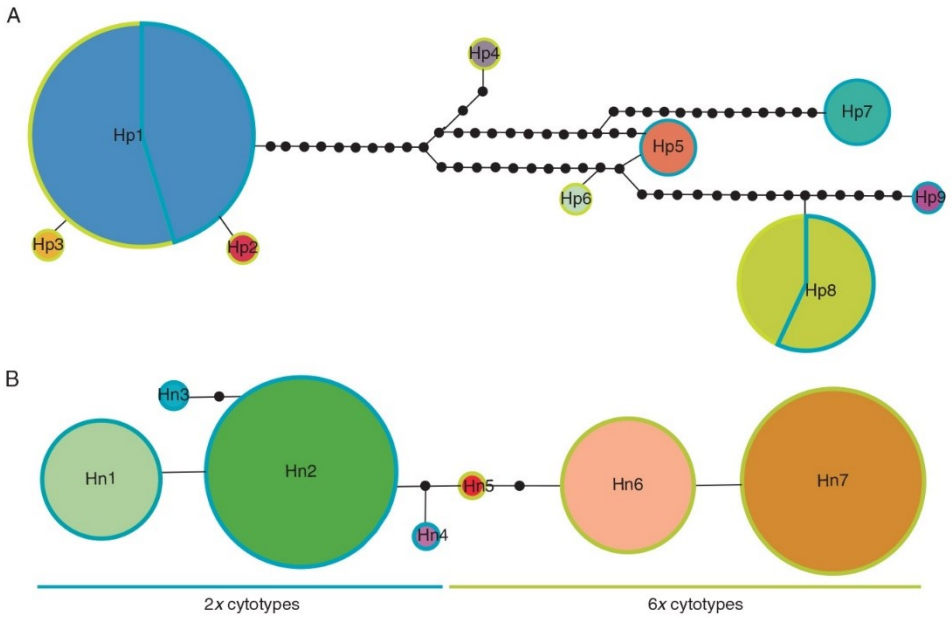


FIG. 4. Statistical parsimony networks inferred using the (A) chloroplast and (B) nuclear sequences by TCS. Black points on connecting lines indicate nucleotide changes. The circle size of the pie charts is proportional to the frequency of haplotypes. Pie charts are surrounded in blue for the diploid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

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

Ecological niche modelling

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

DISCUSSION

Can diploids and hexaploids be considered as distinct evolutionary lineages?

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

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

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

The mismatch distribution analyses are shown in Fig. 5.
Asterisks denote significant differences: $*P < 0.05$, $**P < 0.001$.

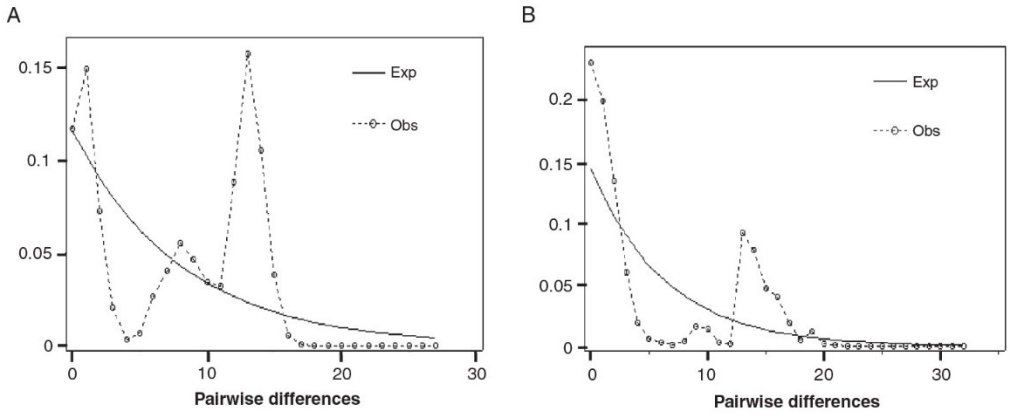


FIG. 5. Demographic analyses for the diploid (A) and hexaploid (B) cytotypes of *Aster amellus* showing the mismatch distribution. The observed (Obs) and expected (Exp) mismatch distributions (left) show the frequencies of pairwise differences.

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

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

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

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

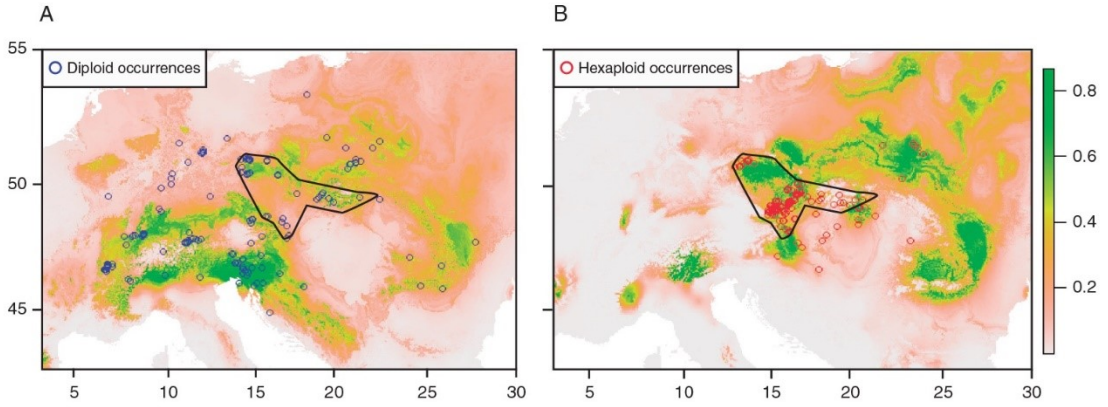


FIG. 6. Geographic projections of the climatic niche model of the (A) diploid and (B) hexaploid cytotypes of *Aster amellus* for current climate. The color scale indicates habitat suitability. The area delimited geometrically in black corresponds to the contact area of the cytotypes. The *x*-axis shows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

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

Multiple-origin vs. single-origin model

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

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

Changes in population size and niche differentiation of the cytotypes

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

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

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

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

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

Conclusions

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

SUPPLEMENTARY DATA

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

ACKNOWLEDGEMENTS

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

LITERATURE CITED

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Study 3: Absence of gene flow between diploids and hexaploids of *Aster amellus* at multiple spatial scales

Heredity

Heredity (2013) 110, 123–130
© 2013 Macmillan Publishers Limited All rights reserved 0018-067X/13
www.nature.com/hdy

ORIGINAL ARTICLE

Absence of gene flow between diploids and hexaploids of *Aster amellus* at multiple spatial scales

Z Münzbergová[1,2], M Šurinová[1,2] and S Castro[3]

The potential for gene exchange across ploidy levels has long been recognized, but only a few studies have explored the rate of gene flow among different cytotypes. In addition, most of the existing knowledge comes from contact zones between diploids and tetraploids. The purpose of this paper was to investigate relationships between diploid and hexaploid individuals within the *Aster amellus* aggregate. *A. amellus* is known to occur in diploid and hexaploid cytotypes in Europe, with a complex contact zone in central Europe. Patterns of genetic diversity were investigated using seven microsatellite loci at three different spatial scales: (1) in the single known mixed-ploidy population; (2) in populations at the contact zone and (3) in a wider range of populations across Europe. The results show clear separation of the cytotypes at all three spatial scales. In addition, analysis of molecular variance strongly supported a model predicting a single origin of the hexaploids, with no or very limited gene flow between the cytotypes. Some hexaploid individuals found in the mixed-ploidy population, however, fell into the diploid cluster. This could suggest recurrent polyploid formation or occasional cross-pollination between cytotypes; however, there are strong post-zygotic breeding barriers between the two cytotypes, making the latter less plausible. Overall, the results suggest that the cytotypes could represent two cryptic species. Nevertheless, their formal separation is difficult as they cannot be distinguished morphologically, occupy very similar habitat conditions and have largely overlapping distribution ranges. These results show that polyploid complexes must be treated with caution as they can hide biological diversity and can have different adaptation potentials, evolving independently.
Heredity (2013) **110,** 123–130; doi:10.1038/hdy.2012.87; published online 21 November 2012

Keywords: autopolyploidy; contact zone; microsatellites; mixed-ploidy population; POLYSAT; STRUCTURE

INTRODUCTION

Genome duplication has long been recognized as a major mechanism of sympatric speciation in flowering plants (Stebbins, 1950; Masterson, 1994; Soltis *et al.*, 2009). As a result, species or species complexes consisting of several independent cytotypes have been documented in many families (for example, Buggs and Pannell, 2007; Halverson *et al.*, 2008; Kloda *et al.*, 2008; Trávníček *et al.*, 2010; Münzbergová *et al.*, 2011).

Understanding the mechanisms of structuring and maintaining these complexes, and their possible adaptations to novel environments requires detailed information on the frequency of gene flow and degree of isolation between the diploid and polyploid entities (Kloda *et al.*, 2008; Symonds *et al.*, 2010). In some cases, strong reproductive barriers may actually lead to the genesis of cryptic species (Soltis *et al.*, 2007), while in other cases there may be substantial gene flow or shared ancestral variation among the cytotypes (Ramsey *et al.*, 2008). Gene flow from diploid to polyploid species could have significant effects on the morphology and ecology of polyploids and affect both their coexistence at the contact zones and possible adaptations to novel environments (Stebbins, 1971; Chapman and Abbott, 2010). Even though the potential for gene exchange across ploidy levels has long been recognized (Stebbins, 1971; Levin, 1975; Chapman and Abbott, 2010), the knowledge on rate of gene flow among different cytotypes in natural systems is still rather limited (for example, Slotte *et al.*, 2008; Chapman and Abbott, 2010; Fehlberg and Ferguson, 2012; Mráz *et al.*, 2012).

Despite its importance, estimating rate of gene flow in polyploid complexes is strongly limited due to difficulties in allele dosage assessment and complex polysomic inheritance in polyploids (Stebbins, 1950, 1971; Clark and Jasieniuk, 2011). Recently, such estimates have been made easier by development of several software programs dealing with these issues (for example, POLYSAT—Clark and Jasieniuk, 2011, STRUCTURE—Falush *et al.*, 2007 and so on).

A. amellus agg. is a widely distributed polymorphic species (Münzbergová *et al.*, 2011). In Europe, this species is known to occur as diploid and hexaploid cytotypes, which are not distinguishable morphologically (Mandáková and Münzbergová, 2008), but form a largely diffuse contact zone across Poland, the Czech Republic, Slovakia, Austria and Romania (Castro *et al.*, 2012). Despite growing in parapatry (except in one population; Castro *et al.* 2012), comparison of habitat requirements of the two cytotypes found no niche separation (Mandáková and Münzbergová, 2006) and each cytotype was able to grow on localities of the other cytotype (Raabová *et al.*, 2008). Previous studies also indicated strong post-pollination breeding barriers between the two cytotypes (Castro *et al.*, 2011).

The existence of diploid and hexaploid cytotypes growing in close proximity without the occurrence of an intermediate tetraploid cytotype and the absence of flowering triploids raises questions on

[1]Department of Botany, Faculty of Science, Charles University, Prague, Czech Republic; [2]Institute of Botany, Academy of Sciences of the Czech Republic, Průhonice, Czech Republic and [3]Faculty of Science and Technology, Department of Life Sciences, CFE, Centre for Functional Ecology, University of Coimbra, Coimbra, Portugal
Correspondence: Dr Z Münzbergová, Department of Botany, Faculty of Science, Charles University, Benátská 2, Prague 12801, Czech Republic.
E-mail: zuzmun@natur.cuni.cz
Received 12 June 2012; revised 4 October 2012; accepted 8 October 2012; published online 21 November 2012

the origin of the higher cytotype, as well as on the relationships between cytotypes. Triploid individuals, as a possible step towards hexaploids, have rarely been detected within the diploid populations and no hexaploid individuals have been detected in populations identified in the past as diploid (Castro *et al.*, 2012). This suggests that the current distribution patterns are a result of secondary contact between cytotypes migrating from the center of their primary origin, most likely in south-eastern Europe or Asia (Münzbergová *et al.*, 2011). However, additional primary contact zones could theoretically occur in other areas throughout the distribution range (Castro *et al.*, 2012). Genetic relationships between diploid and hexaploid populations from the Czech Republic studied using allozymes indicated significant genetic separation between the two cytotypes (Mandáková and Münzbergová, 2008). The marker is, however, very conservative and low levels of genetic exchange between the cytotypes could thus remain undetected. Also, the previous sampling was spatially limited and did not include the single mixed-ploidy population detected so far.

The aim of this study was to assess relationships between the diploid and hexaploid cytotypes of *A. amellus* using highly variable microsatellite markers at three spatial scales: (1) within the single known mixed-ploidy population; (2) between pairs of diploid and hexaploid populations at the contact zone; and (3) between diploid and hexaploid individuals distributed across Europe. We used several alternative approaches to analyze the data and assessed if the different approaches provided comparable results and affected interpretation of patterns of gene flow among cytotypes.

MATERIALS AND METHODS

Study species

A. amellus L. (Asteraceae) is a polymorphic perennial herb growing in open xerothermic habitats. In Europe, its distribution area ranges from northern France to Lithuania, reaching Italy and Macedonia in the south, to the Black Sea, Caucasus and west Siberia outside Europe (Münzbergová *et al.*, 2011). It has been suggested that *A. amellus* comprises three cytotypes: diploids ($2n = 2x = 18$), tetraploids ($2n = 4x = 36$), and hexaploids ($2n = 6x = 54$) (Münzbergová *et al.*, 2011). However, only diploid and hexaploid cytotypes have been confirmed both in a large-scale screening of populations across Europe and in a literature survey (Münzbergová *et al.*, 2011; Castro *et al.*, 2012). Previous studies also show that the diploid and hexaploid cytotypes are not morphologically distinguishable (Mandáková and Münzbergová, 2008). In addition, we used allozyme analyses to compare the two cytotypes at the contact zone in the Czech Republic. Analyses of the data (presence/absence of the alleles) using principal component analyses (PCA), as well as a dendrogram based on Nei's genetic distances between populations constructed using Unweighted Pair Group Method with Arithmetic Mean (UPGMA) method of clustering clearly separated the two cytotypes. There were, however, very few unique alleles in each cytotype and all of these had very-low frequency. In addition, high frequency of unbalanced heterozygotes was found in the hexaploid banding patterns (Mandáková and Münzbergová, 2008). All this was used to conclude an autopolyploid origin for the hexaploids and that the contact zone is likely of secondary origin. The study of reproductive barriers between *A. amellus* cytotypes indicate that the absence of hybrids (tetraploids) in the field is mostly due to spatial segregation of diploids and hexaploids (Mandáková and Münzbergová, 2006; Castro *et al.*, 2012) and to strong post-pollination barriers (Castro *et al.*, 2011).

While a lot of previous knowledge has already been accumulated on the system, all the existing knowledge on genetic relationships between populations comes only from a small part of the contact zone, the Czech Republic, and was assessed using a very conservative marker, the allozymes.

Study populations

Genetic relationships between plants at a small spatial scale were studied in the Strebensdorf population (Austria), the only mixed-ploidy population that has been detected so far (Castro *et al.*, 2012, Figure 1, Supplementary Information). In this population, every adult plant (both reproductive and vegetative plants) was labeled, its position recorded in a rectangular coordinate system, its DNA ploidy level measured using flow cytometry (Castro *et al.*, 2012) and a sample for DNA analyses collected and stored in silica gel. In total, the sampling comprised 7 diploid and 61 hexaploid individuals.

To detect genetic relationships at a larger spatial scale, nine pairs of diploid and hexaploid populations growing in close proximity (between 8 and 50 km apart, mean 14.4 km) and located across the contact zone in Austria, the Czech Republic and Slovakia were selected (Figure 1, Supplementary Information). Five plants were sampled from each population.

To analyze the data at the largest spatial scale, we selected 22 diploid and 22 hexaploid populations covering most of the European range (and a small part of Asia) of the species. We selected three plants per population for this analysis (Figure 1, Supplementary Information).

At all three spatial scales, two samples per plant were collected, one for ploidy assessment and the other for DNA analyses.

Microsatellite analyses

Total genomic DNA was extracted from dehydrated leaves and stored in silica gel using a sorbitol extraction method (Štorchová *et al.*, 2000). Seven microsatellite loci were amplified: *Aam*.F58, *Aam*.A12, *Aam*.H231, *Aam*.B3, *Aam*.A415, *Aam*.J15 and *Aam*.G431 using specific primers for each region (Mayor and Naciri, 2007). Multiplex PCR reactions (Multiplex Mix) were designed combining microsatellites A12, A415, B3, F58 (Multiplex Mix 1) and J15, H231, G431 (Multiplex Mix 2). QIAGEN Multiplex PCR Master Mix (Qiagen, Hilden, Germany) containing HotStartTaq DNA polymerase (Lucigen, Middleton, WI, USA), a multiplex PCR buffer at 6 mM $MgCl_2$, deoxynucleotide triphosphates and a factor MP (Qiagen) that improves annealing and elongation were used. Initial Multiplex PCR reactions were designed for diploid and hexaploid individuals simultaneously. Afterwards, primer concentrations were adjusted separately for each cytotype for optimal and comparable electrophoretograms. Specifically, primer concentrations had to be increased for hexaploids because of the three times large genome, and thus higher total number of microsatellite sites.

Final Multiplex mix 1 PCR amplifications for diploids contained 5 µl QIAGEN Multiplex PCR Master Mix, 0.25 µl of A12, A415 and B3 primers and 0.35 µl of F58 primers (10 µM each in initial volume), 1.6 µl H_2O and 24 ng of DNA dissolved in 1.2 µl TE buffer. Final Multiplex mix 2 PCR amplifications for diploids contained 5 µl QIAGEN Multiplex PCR Master Mix, 0.25 µl of J15 and H231 primers and 0.35 µl of G431 primers (10 µM each in initial volume), 2.1 µl H_2O and 24 ng of DNA dissolved in 1.2 µl TE buffer. Final Multiplex mix 1 PCR amplifications for hexaploids contained 5 µl QIAGEN Multiplex PCR Master Mix, 0.25 µl of B3 primers, 0.35 µl of A12, A415 primers and 0.5 µl of F58 primers (10 µM each in initial volume), 0.9 µl H_2O and 24 ng of DNA dissolved in 1.2 µl TE buffer. Final Multiplex mix 2 PCR amplifications for hexaploids contained 5 µl QIAGEN Multiplex PCR Master Mix, 0.35 µl of J15, H231, G431 primers (10 µM each in initial volume), 1.7 µl H_2O and 24 ng DNA dissolved in 1.2 µl TE buffer were also added. An initial denaturation step at 95 °C for 15 min was followed by 35 cycles of denaturation (94 °C for 30 s), annealing (58 °C for 90 s) and extension (72 °C for 60 s) steps, and a final extension step at 60 °C for 30 min. A 1 µl aliquot of the PCR product was mixed with 11 µl of a 120: 1 solution of formamide and size standard GeneScan 500 LIZ (Life Technologies, Foster City, CA, USA). Fragment lengths were determined by capillary gel electrophoresis with an ABI 3130 Genetic Analyzer using Gene Mapper 4.0 (Life Technologies) and the peaks were scored manually.

To check the reproducibility of our microsatellite analyses, we selected three diploid and three hexaploid individuals, and repeated the microsatellite analyses of these in three independent runs for each individual. The results indicated 100% match in the results, suggesting that our microsatellite analyses are highly reproducible.

Data analyses

When scoring the data, for the hexaploids it was not possible to distinguish the exact number of copies for a given allele. Due to this, we worked only with the presence and absence of alleles.

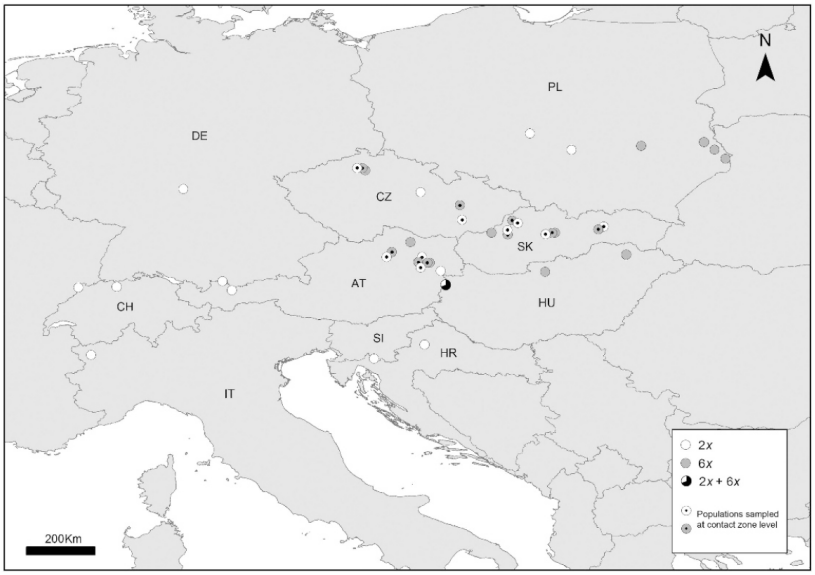


Figure 1 Location of the populations of *Aster amellus* studied. Cytotypes are represented by different colors: white circles-diploid (2*x*); grey circles-hexaploid (6*x*). The mixed population is represented by a pie diagram representing cytotype proportions (white-2*x*; black-6*x*). All locations were used in the large-scale spatial analyses; locations with dots represent the populations used in the intermediate spatial scale (contact zone). The population from Turkey (2x) is not displayed on the map. Country codes: AT, Austria; CH, Switzerland; CZ, Czech Republic; DE, Germany; HR, Croatia; HU, Hungary; IT, Italy; PL, Poland; SI, Slovenia; SK, Slovakia.

The microsatellite results were used to calculate several genetic summary statistics for each of the three spatial scales studied, including total the number of alleles, observed (H_O) and expected heterozygosity (H_E), inbreeding coefficient (F_{IS}), and distribution of genetic diversity within and between cytotypes. The values of H_E and F_{IS} were calculated using SPAGeDi 1.2 (Hardy and Vekemans, 2002). s.e. of F_{IS} was obtained by jackknifing over loci.

Because of the unknown dosage of single alleles in the hexaploids, the H_E values for hexaploids are values of expected proportion of individuals bearing more than one allele and thus correspond to zygotic heterozygosity. This applies also for H_O values. Total gene diversity (H_T) was estimated as the H_E value for the whole data set, and the mean genetic diversity within populations (H_S) was calculated as the mean of the mean H_E value of all diploids in the data set, and mean H_E of all hexaploids in the data set. Genetic diversity among populations (D_{ST}) was calculated as $D_{ST} = H_T — H_S$. In case of the mixed population, the D_{ST} value represented differentiation between cytotypes within the single population.

To analyze relationships between the two cytotypes, we performed PCA using POLYSAT software (Clark and Jasieniuk, 2011). POLYSAT, running under the R platform, is a software program for analyzing autopolyploid and allopolyploid microsatellite data. It handles genotype data of any ploidy, including populations of mixed-ploidy, and assumes that the allele copy number is always ambiguous in partial heterozygotes. A pair-wise distance matrix between all samples was calculated using Bruvo distances (Bruvo *et al.*, 2004), which takes distances between microsatellite alleles into account without the knowledge of allele copy number (Clark and Jasieniuk, 2011). The distance matrix was used as a correlation matrix needed for visualizing the results using PCA).

Differentiation between pairs of populations as a function of their geographic distance was assessed using two different measures of pair-wise between-population differentiation: the Bruvo distance calculated above and the *Rho* statistic (Ronfort *et al.*, 1998). *Rho* is an inter class relatedness coefficient permitting comparison among ploidy levels and it was calculated using SPAGeDi at the intermediate and large spatial scale (it is not defined in the case of a single population). For Bruvo distances, we used the original data at the smallest spatial scale and population means at the intermediate and large spatial scales. In both cases, we calculated distances between diploid population pairs, hexaploid population pairs and inter-cytotype pairs. Isolation by distance was inspected through the regression of those values on the logarithm of geographic distance according to Hardy and Vekemans (2001). Statistical significance was tested using a Mantel test with 999 permutations. The *Rho* statistics was also used to compare the overall differentiation between populations within cytotypes and between cytotypes. In the case of the mixed population, the *Rho* value represented differentiation between cytotypes within the single population. Values of standard error of *Rho* were obtained by jackknifing over loci.

We also used STRUCTURE version 2.2 (Pritchard *et al.*, 2000) to analyze the data. STRUCTURE is a Bayesian population assignment and inference software program widely used to detect population genetic structure (Kaeuffer *et al.*, 2007). It implements algorithms accounting for genotypic uncertainty arising from copy number variation when the data includes polyploid cytotypes. Under the single-origin model, we would expect the *K* clusters identified by STRUCTURE to correspond to cytotypes, whereas under the multiple origins model we would expect clusters corresponding to sites. For each data set, we ran STRUCTURE using the admixture model and varying *K* values (up to the number of sub-populations) ranging from one to five at the small spatial scale and 10 at the medium and large spatial scales. Each run consisted of 100 000 burn-in iterations and 100 000 data collection iterations. Each value of *K* was evaluated using 20 independent Markov chain Monte Carlo replicates. We inferred the number of clusters following Pritchard *et al.* (2000), with posterior probabilities of *K* calculated assuming uniform priors on *K* and using for each *K* the maximum value of the probability of the data given *K*, $\ln \Pr(X|K)$, obtained over Markov chain Monte Carlo replicates (Halverson *et al.*, 2008).

At the intermediate spatial scale we also tested two contrasting models explaining genetic relationships among cytotypes using analysis of molecular variance (AMOVA) (Excoffier *et al.*, 1992), as implemented in GenAlEx 6 (Peakall and Smouse, 2006) by following the method described in Halverson *et al.* (2008). Specifically, we compared the single-origin model and a model with multiple origins of the cytotypes. The single-origin model proposes that

the hexaploid cytotype arose only once from the diploid cytotype, and each cytotype currently represents a monophyletic lineage. This model also necessarily proposes that the origins of the polyploid cytotypes took place long enough ago for the resultant cytotypes to have spread across their modern range (Halverson *et al*., 2008). We tested this model by a hierarchical AMOVA model in which genotypes are nested within localities and localities within cytotypes. Alternatively, the strictest multiple origins model proposes that polyploid cytotypes have arisen repeatedly and independently (and perhaps recently) at each geographic location (Halverson *et al*., 2008). This hypothesis was tested using an AMOVA analysis in which genotypes were nested within cytotypes and these were nested within regions (pairs of nearby diploid and hexaploid populations). The preferred AMOVA model is that in which the higher stratum (cytotypes for the single-origin model or region for the multiple origins model) explains a greater component of total genetic variation (Halverson *et al*., 2008).

RESULTS

The total number of alleles found across cytotypes ranged between 104 in the mixed-ploidy population and 179 at the largest spatial scale (Table 1). In all the three data sets, the number of alleles found in hexaploids exceeded the number found in the diploids (Table 1). The total number of alleles across cytotypes did not strongly exceed the number of alleles found in hexaploids, indicating relatively large overlap in identity of alleles between cytotypes. This large overlap in the identity of alleles was also confirmed by the relatively small D_{ST} values, which indicates relatively low differentiation between cytotypes when compared with total genetic diversity (H_T). Similarly, also the D_{ST} within populations was quite small compared with total genetic diversity within cytotypes (H_T) indicating large similarity between populations within cytotypes (Table 1).

In spite of the relatively large overlap between cytotypes as indicated by D_{ST} values, the two cytotypes are significantly differentiated at all scales as indicated by *Rho* coefficient, which is significantly deviating from 0 in all cases. In addition, significant differentiation between cytotypes is also visible from AMOVA analyses (Table 1).

In all cases, the inbreeding coefficient (F_{IS} value) was negative. The values, however, significantly deviating from 0 only for the hexaploid cytotype, indicating strong heterozygote excess in the hexaploids (Table 1).

Mixed-ploidy population

PCA analyses performed by POLYSAT clearly separated the two cytotypes into separate clusters along the first axis (Figure 2a). The first axis separating the cytotypes explained 31.9% of the total variation in the data set, while the second axis accounting for variation between individual plants within cytotypes explained only 10.9% of the total variation.

Similarly to the PCA, the analysis performed in STRUCTURE identified two clusters of individuals ($K=2$ having a posterior probability greater than 0.999 relative to other tested values of K; Table 2). All diploid individuals had >93% probability of falling into one of the clusters (mean ± s.d., 94.1 ± 0.5%). The assignment of all hexaploids into the second cluster was less certain (91.7 ± 10.8%). Fifty-six hexaploid individuals had >90% probability of falling into the second cluster. However, the remaining five individuals had 88, 87, 67, 47 and 28% probability of falling into the cluster containing mostly hexaploids. Therefore, two of these had <50% chance to be a member of the cluster containing the diploids (Figure 3a). These five hexaploid individuals also can be seen as outliers in the PCA graph and represented the five hexaploid points closest to the diploid cluster (Figure 2a).

Table 1 Summary characteristics describing genetic diversity assessed using microsatellites in the three data sets, in each case separately for diploids, hexaploids and for the total data set

	Mixed-ploidy population			*Contact zone*			*Europe*		
	2x	*6x*	*Combined*	*2x*	*6x*	*Combined*	*2x*	*6x*	*Combined*
TA	28	96	104	90	129	160	98	150	179
F_{IS} (± s.e.)	−0.045 ± 0.11	−0.232 ± 0.12*	−0.0371 ± 0.02s	−0.080 ± −0.11s	−0.099 ± 0.02*	−0.002 ± 0.05	−0.035 ± 0.07	−0.154 ± 0.03*	−0.013 ± 0.003
H_E	0.642	0.762	0.786	0.816	0.834	0.886	0.803	0.837	0.882
H_O	0.671	0.939	0.815	0.881	0.917	0.887	0.83	0.97	0.89
H_T	NA	NA	0.786	0.816	0.834	0.886	0.803	0.837	0.882
H_S	NA	NA	0.702	0.753	0.827	0.825	0.697	0.815	0.820
D_{ST}	NA	NA	0.084	0.063	0.007	0.061	0.106	0.022	0.062
Rho (± s.e.)	NA	NA	0.511 ± 0.07*	0.182 ± 0.03*	0.115 ± 0.03*	0.292 ± 0.05*	0.257 ± 0.03*	0.177 ± 0.02*	0.347 ± 0.03*
AMOVA									
Among regions	NA	NA	NA	NA	NA	0%	NA	NA	NA
Among cytotypes	NA	NA	31%*	NA	NA	26%*	NA	NA	22%*
Among populations	NA	NA	NA	17%*	14%*	11%*	24%*	18%*	16%*
Within populations	NA	NA	69%*	83%*	86%*	63%*	76%*	82%*	62%*

Abbreviations: AMOVA, analysis of molecular variance; D_{ST}, genetic diversity among populations; F_{IS}, inbreeding coefficient; H_E, expected heterozygosity; H_O, observed heterozygosity; H_S, mean genetic diversity within populations; H_T, total gene diversity; NA, not applicable; TA, total number of alleles.
Values of F_{IS} and *Rho*, significantly deviating from 0, are marked by *. Also shown are the proportions of genetic variation in the data set at each hierarchical level as assessed by AMOVA; again those significantly deviating from 0 are marked by *. In the mixed-ploidy population, the 'populations' in the table are groups of diploid and group of hexaploid individuals.

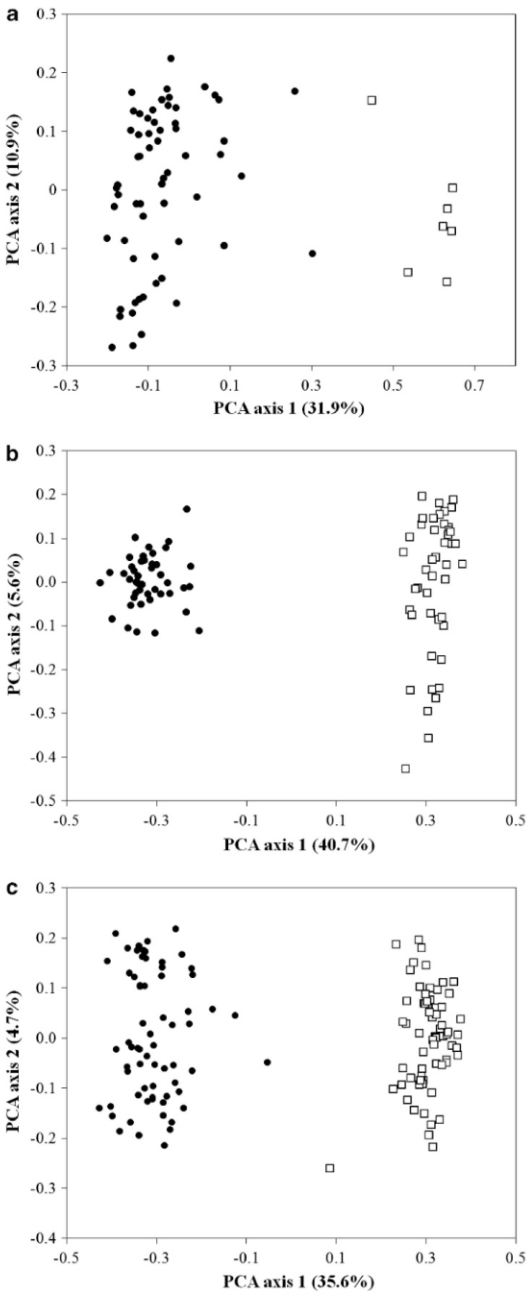


Figure 2 PCAs using Bruvo distances performed in POLYSAT. (**a**) Mixed-ploidy population. (**b**) Contact zone between cytotypes. (**c**) Large spatial scale across Europe. Open squares indicate diploids, filled circles represent hexaploids. % of variance explained by each axis is provided within the figure.

There was significant isolation by distance among the hexaploid individuals, but not among the diploid individuals or between the diploid and hexaploid individual pairs based on Bruvo distances (Table 3).

Multiple populations across the contact zone

Similarly to what occurred with the mixed-ploidy population, PCA performed using POLYSAT at the contact zone clearly separated the two cytotypes (Figure 2b). The first axis separating the two cytotypes explained 40.7% of the total variation in the data set. In contrast, the second axis representing variation between populations explained only 5.6% of the total variation in the data set. The variation between diploid populations was much higher than the variation between hexaploid populations.

STRUCTURE analysis also identified two clusters of individuals ($K=2$ having a posterior probability >0.99 relative to other tested values of K; Table 2), clearly separating the two cytotypes. All diploid individuals had $>99\pm1.5\%$ probability to fall into one of the clusters and all the hexaploids had more than $99\pm0.1\%$ to fall into the other cluster (Figure 3b).

There was significant isolation by distance among hexaploid populations as assessed by the *Rho* coefficient and marginally significant isolation by distance as assessed by Bruvo distance (Table 3). There was no significant isolation by distance between diploid population pairs or between diploid and hexaploid population pairs based on either of the two distance measures (Table 3).

The AMOVA analyses comparing the single-origin polyploidy model and strict multiple origins polyploidy model clearly indicated that the single-origin model is superior to the multiple origin model. In the single-origin model, cytotype explained 26% of the total genetic variation and was highly significant. In contrast, the region in the multiple origins model explained 0% of the total variation and was not significant (Table 4). The variation between populations accounted for 11% of the variation in the data set (Table 1).

Populations sampled across Europe

PCA analyses performed by POLYSAT at the largest spatial scale confirmed the clear separation between cytotypes (Figure 2c). The first axis separating the cytotypes explained 35.6% and the second axis capturing variation among populations explained 4.7% of the total variation in the data set. In contrast to the data at the contact zone, variation between populations in the diploid cytotype was comparable to variation between populations in the hexaploid cytotype.

STRUCTURE analysis at the largest spatial scale identified, in contrast to the two smaller scales, six clusters of individuals ($K=6$ having a posterior probability of >0.99 relative to other tested values of K; Table 2). Two spatial clusters were identified as the second most likely option. One of the clusters clearly corresponded to the diploids, while the other five clusters corresponded to the hexaploids. Most of the hexaploid individuals, however, appeared to be a mixture of all five types. (Figure 3c).

At the largest spatial scale there was significant isolation by distance only between hexaploid population pairs, detected using Bruvo distances (Table 3).

The AMOVA analyses partitioning variation among population and among cytotypes at the largest scale indicated that variation among cytotypes is slightly higher (22%) than variation among populations (16%, Table 1).

DISCUSSION

The results obtained at the three spatial scales studied indicate relatively strong genetic separation between diploid and hexaploid cytotypes of *A. amellus*. This contrasts with many previous studies on diploid–polyploid systems, in which either frequent gene flow or broad shared ancestral polymorphisms among cytotypes has been reported (for example, Ramsey *et al.*, 2008; Fehlberg and Ferguson,

Table 2 Results of STRUCTURE analyses showing log-likelihood estimation of cluster number for $K = 1$–5–(10)

Data set	K	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Mixed-ploidy population	ln Pr(X|K)	−3511	**−3393**	−3817	−3632	−3513					
	Δ ln-likelihood	−118	**0**	−424	−239	−120					
	Posterior Pr(K)	<0.001	>**0.999**	<0.001	<0.001	<0.001					
Contact zone	ln Pr(X|K)	−4479	**−4081**	−4131	−4185	−4200	−4284	−4378	−4293	−4328	−4412
	Δ ln-likelihood	−398	**0**	−49	−103	−119	−202	−297	−212	−247	−331
	Posterior Pr(K)	<0.001	>**0.999**	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Europe	ln Pr(X|K)	−6594	−5979	−5985	−6166	−6154	**−5896**	−6363	−6270	−6195	−6197
	Δ ln-likelihood	−698	−83	−89	−270	−258	**0**	−467	−374	−299	−301
	Posterior Pr(K)	<0.001	<0.001	<0.001	<0.001	<0.001	>**0.999**	<0.001	<0.001	<0.001	<0.001

The results are shown for the small (mixed-ploidy population), intermediate (contact zone) and large (Europe) spatial scales. The favored *K* values are in bold

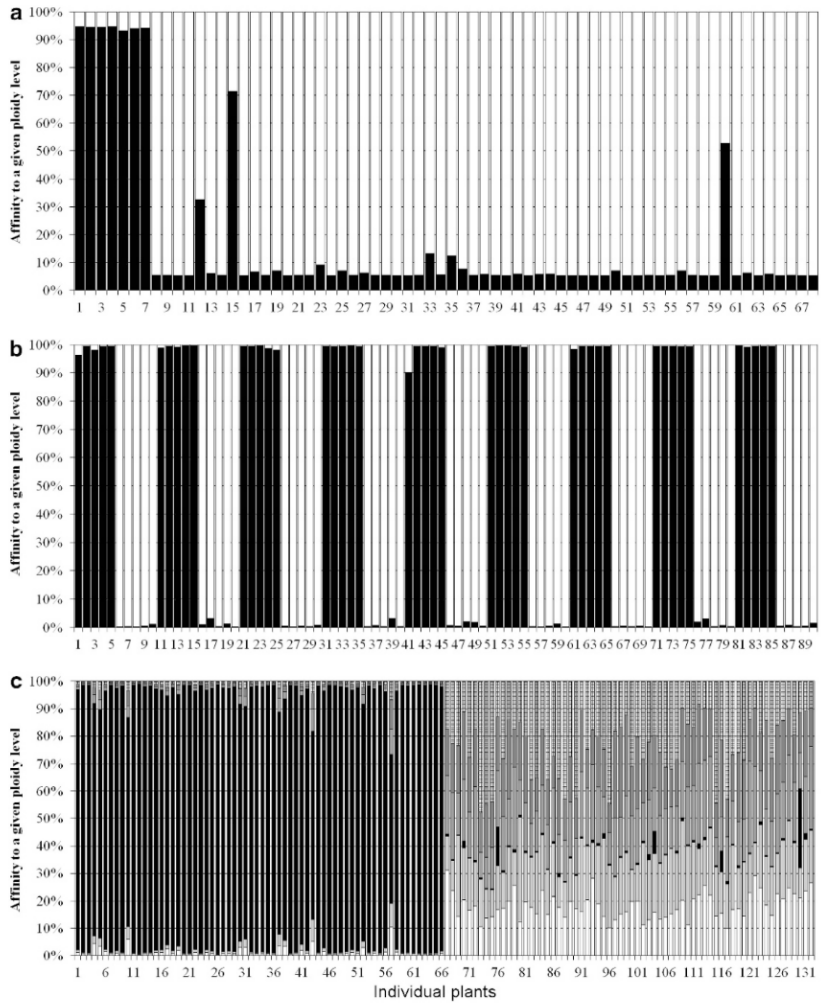


Figure 3 Individual assignment based on Bayesian genotypic cluster analysis performed using STRUCTURE. Histograms demonstrate the proportion of each individual's genome that can be assigned to each of the clusters. (**a**) Mixed-ploidy population (individuals 1–7 are diploid, the remaining are hexaploid) at $K = 2$. (**b**) Contact zone between cytotypes at $K = 2$ (individuals are sorted by pairs of populations, the diploid population within each pair goes first; that is individuals 1–5 are diploids of pair 1, individuals 6–10 are hexaploids of pair 1, individuals 11–15 are diploids of pair 2). (**c**) Large spatial scale across Europe at $K = 6$ (Individuals 1–66 are diploid, 67–132 are hexaploid).

Table 3 Correlation between pair-wise distances between populations measured using *Rho* coefficients and Bruvo distances, and the logarithm of the geographical distance

	Mixed-ploidy population			*Contact zone*			*Europe*		
	2x	*6x*	*2x–6x*	*2x*	*6x*	*2x–6x*	*2x*	*6x*	*2x–6x*
Rho									
r	NA	NA	NA	0.04	**0.46**	0.04	0.06	0.15	0.03
p	NA	NA	NA	0.45	**0.04**	0.74	0.30	0.13	0.57
Bruvo									
r	0.18	**0.17**	0.08	0.16	0.42	0.07	0.12	**0.20**	0.19
p	0.34	**0.01**	0.14	0.24	0.07	0.55	0.17	**0.02**	0.42

Abbreviation: NA, not applicable.
The calculations were done for diploid population pairs, hexaploid population pairs and diploid-hexaploid population pairs. Values significant at $P \leqslant 0.05$ are in bold.

Table 4 AMOVA analyses at the intermediate spatial scale testing the (A) single-origin polyploidy model and (B) strict multiple origins polyploidy model

Source of variation	*d.f.*	*Sum of sq.*	*Percentage of variance*	f	P
(A) Single-origin polyploidy model					
Among cytotypes	1	206.267	26	0.266	**0.001**
Among localities w/in cytotypes	16	293.733	11	0.148	**0.001**
Individuals w/in localities	72	708.200	63	0.374	**0.001**
(B) Strict multiple origins polyploid model					
Among regions	8	150.100	0	−0.148	1.000
Among cytotypes w/in regions	9	349.900	37	0.371	**0.001**
Individuals w/in localities	72	708.200	63	0.279	**0.001**

Values significant at $P \leqslant 0.05$ are in bold.

2012). Thus, our results suggest that the two cytotypes are largely evolving independently (Soltis and Soltis, 1999) and could theoretically be viewed as two cryptic species (Soltis *et al.*, 2007). Consequently, the two cytotypes are likely to follow different evolutionary trajectories and accumulate various divergent adaptations. In spite of this, their formal separation is difficult as they cannot be distinguished morphologically (Mandáková and Münzbergová, 2008), occupy very similar habitat conditions (Mandáková and Münzbergová, 2008) and have largely overlapping distribution ranges (Castro *et al.*, 2012).

Strong separation among cytotypes was observed at all the three spatial scales. The relatively consistent results across the three spatial scale contrast with our expectations, based on the previous studies (for example, Hardy and Vekemans, 2001), that the strength of the separation between cytotypes will depend on spatial scale. This pattern, it means strong separation independent of scale, is probably due to the strong breeding barriers between the cytotypes (Castro *et al.*, 2011) and very limited dispersal ability of the species (Münzbergová *et al.*, 2011).

The test of the single versus multiple origin models at the intermediate spatial scale strongly supported the single-origin model, indicating that even at this scale, there is no or very limited gene flow between the cytotypes. On the basis of this we predict that the studied contact zone is of secondary origin, similarly to many contact zones between different cytotypes described in the literature (for example, Petit *et al.*, 1999; Hardy and Vekemans, 2001).

In spite of the overwhelming evidence that the two cytotypes of *A. amellus* are separated, and that hexaploids are of single origin, there are also several lines of evidence suggesting that some hexaploid populations may have arisen *in situ* and thus that multiple origin could also occur in *A. amellus*. Specifically, occasional recurrent formation of hexaploid individuals in a diploid population followed by rapid reproductive isolation could explain the pattern. This is most clearly seen for the mixed population, in which several hexaploid individuals fell with high probability into a diploid cluster (as suggested by analyses in STRUCTURE and seen also in the PCA). Thus, it can be speculated that the hexaploid individuals might have originated *in situ* from the diploids and become established. The hexaploids could subsequently have become more abundant because of higher reproductive success in comparison with diploids (Castro *et al.*, 2012) and the diploid cytotype (currently occurring as only seven mature plants at one edge of the population) is currently being excluded from the population (minority cytotype exclusion; Levin, 1975; Rodriguez, 1996). If this theory is correct, it could be confirmed by revisiting the population in several years and identifying changes in the proportion of the cytotypes. A similar scenario has been observed, for example, in *Vicia cracca*, where the diploid cytotype completely disappeared from several tetraploid populations (Trávníček *et al.*, 2010). Alternatively, this pattern could also be a result of occasional cross-pollination, followed by introgression. Although such an explanation has been suggested repeatedly for other diploid–polyploid systems (for example, Slotte *et al.*, 2006), we think it is quite unlikely in this system, due to strong post-pollination breeding barriers between the two cytotypes (Castro *et al.*, 2011).

Multiple origins of higher ploidy levels has also been proposed repeatedly in various other polyploid aggregates (for example, Hardy and Vekemans, 2001; Segraves *et al.*, 2002; Symonds *et al.*, 2010) and multiple origins are commonly predicted even in secondary contact zones, such as the one observed in *A. amellus*. This suggests that repeated origins of a cytotype may be an important mechanism of diversification in many polyploid lineages, including *A. amellus*.

Isolation by distance was detected in the hexaploid cytotype at all the three scales but not for the diploid cytotypes or between cytotype pairs. This pattern is in agreement with theoretical expectations for polyploids, as hexaploid populations should experience lower genetic drift pressures due to polysomic inheritance and thus should be less divergent (for example, Moody *et al.*, 1993; Ronfort *et al.*, 1998, Mahy *et al.*, 2000; Hardy and Vekemans, 2001). Our conclusion of lower differentiation between hexaploid populations compared with diploids is also apparent from lower D_{ST} values among the hexaploid populations. In contrast to our expectations, isolation by distance was similar across the three data sets and as above, this might be considered as the consequence of strong breeding barriers between the cytotypes (Castro *et al.*, 2011) and very limited dispersal ability of the species (Münzbergová *et al.*, 2011).

Overall, our results indicate that the two cytotypes represent two largely independent entities that could be viewed as two cryptic species, with the hexaploid being of autopolyploid origin. The relationships detected among the two cytotypes are largely comparable between the three spatial scales and between different techniques

used to analyze the data. This suggests that our conclusions on the system are quite general and it can be expected that these two cytotype are undergoing largely independent evolutionary trajectories.

More generally, the results indicate that polyploidy can cause genetic divergence between populations of the same species, even in situations when the different cytotypes are morphologically very similar, have overlapped distribution ranges and similar habitat requirements. No or very-low gene flow between the cytotypes gives them a high potential for adaptations to novel environments. Thanks to this, the polyploid cytotypes may gain a large variation of novel properties that may prove as advantageous pre-adaptations in cases the species is transported to novel areas and thus increases its potential for colonization.

DATA ARCHIVING

The data are available from the authors at zuzmun@natur.cuni.cz.

CONFLICT OF INTEREST

The authors declare no conflict of interests.

ACKNOWLEDGEMENTS

This study was supported by GAČR P504/10/0456, partly by project no. RVO 67985939 and by FCT and European Social Fund with the fellowship FCT/BPD/41200/2007 to SC. We thank two anonymous reviewers, Barbara Mable and the Heredity Editor for useful comments on the previous version of the manuscript.

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.

Buggs RJA, Pannell JR (2007). Ecological differentiation and diploid superiority across a moving ploidy contact zone. *Evolution* **61**: 125–140.

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

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

Chapman MA, Abbott RJ (2010). Introgression of fitness genes across a ploidy barrier. *New Phytol* **186**: 63–71.

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

Excoffier L, Smouse P, Quattro J (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**: 479–491.

Falush D, Stephens M, Pritchard JK (2007). Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol Ecol Notes* **7**: 574–578.

Fehlberg SD, Ferguson CJ (2012). Intraspecific cytotypic variation and complicated genetic structure in the *Phlox amabilis*-P. *woodhousei* (Polemoniaceae) complex. *Am J Bot* **99**: 865–874.

Halverson K, Heard SB, Nason JD, Stireman 3rd JO (2008). Differential attack on diploid, tetraploid, and hexaploid *Solidago altissima* L. by five insect gallmakers. *Oecologia* **154**: 755–761.

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.

Kaeuffer R, Réale D, Coltman DW, Pontier D (2007). Detecting population structure using STRUCTURE software: effect of background linkage disequilibrium. *Heredity* **99**: 374–80.

Kloda JM, Dean PD, Maddren C, MacDonald DW, Mayes S (2008). Using principle component analysis to compare genetic diversity across polyploidy levels within plant complexes: an example from British Restharrows (*Ononis spinosa* and *Ononis repens*). *Heredity* **100**: 253–260.

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

Mahy G, Bruederle LP, Connors B, Van Hofwegen M, Vorsa N (2000). Allozyme evidence for genetic autopolyploidy and high genetic diversity in tetraploid cranberry, *Vaccinium oxycoccos* (Ericaceae). *Am J Bot* **87**: 1882–1889.

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

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

Masterson J (1994). Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. *Science* **264**: 421–424.

Mayor R, Naciri Y (2007). Identification and characterization of eight microsatellite loci in *Aster amellus* L. (Asteraceae). *Mol Ecol Notes* **7**: 233–235.

Moody ME, Mueller LD, Soltis DE (1993). Genetic variation and random drift in autotetraploid populations. *Genetics* **134**: 649–657.

Mráz P, Spaniel S, Keller A, Bowmann G, Farkas A, Singliarova B *et al.* (2012). 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.

Münzbergová Z, Raabová J, Castro S, Pánková H (2011). Biological flora of central Europe: *Aster amellus* L. (Asteraceae). *Persp Plant Ecol, Evol Syst* **13**: 151–162.

Peakall R, Smouse PE (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* **6**: 288–295.

Petit C, Bretagnolle F, Felber F (1999). Evolutionary consequences of diploid-polyploid hybrid zones in wild species. *Trends Ecol. Evol* **14**: 306–311.

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

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

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

Rodriguez DJ (1996). A model for the establishment of polyploidy in plants. *Am Nat* **147**: 33–46.

Ronfort J, Jenczewskia E, Bataillona T, Roussetb F (1998). Analysis of population structure in autotetraploid species. *Genetics* **150**: 921–930.

Segraves KA, Thompson JN, Soltis PS, Soltis DE (2002). Multiple origins of polyploidy and the geographic structure of *Heuchera grossulariifolia*. *Mol Ecol* **8**: 253–262.

Slotte T, Ceplitis A, Neuffer B, Hurka H, Lascoux M (2006). Intrageneric phylogeny of *Capsella* (Brassicaceae) and the origin of the tetraploid *C. bursa-pastoris* based on chloroplast and nuclear DNA sequences. *Am J Bot* **93**: 1714–1724.

Slotte T, Huang H, Lascoux M, Ceplitis A (2008). Polyploid speciation did not confer instant reproductive isolation in *Capsella* (Brassicaceae). *Mol Biol Evol* **25**: 1472–1481.

Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng CF *et al.* (2009). Polyploidy and angiosperm diversification. *Am J Bot* **96**: 336–348.

Soltis DE, Soltis PS (1999). Polyploidy: recurrent formation and genome evolution. *Trends Ecol Evol* **14**: 348–352.

Soltis DE, Soltis PS, Schemske DW, Hancock JF, Thompson JN, Husband BC *et al.* (2007). Autopolyploidy in angiosperms: have we grossly underestimated the number of species? *Taxon* **56**: 13–30.

Stebbins GJ (1950). *Variation and evolution in plants*. Columbia University Press: New York.

Stebbins GJ (1971). *Chromosomal Evolution in Higher Plants*. Addison-Wesley: London.

Symonds VV, Soltis PS, Soltis DE (2010). Dynamics of polyploid formation in *Tragopogon* (Asteraceae): Recurrent formation, gene flow, and population structure. *Evolution* **64**: 1984–2003.

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.

Štorchová H, Hrdličková R, Chrtek J, Tetera M, Fitze D, Fehrer J (2000). An improved method of DNA isolation from plants collected in the field and conserved in saturated NaCl/CTAB solution. *Taxon* **49**: 79–84.

Supplementary Information accompanies the paper on Heredity website (http://www.nature.com/hdy)

Study 4: Temperature and precipitation, but not geographic distance, explain genetic relatedness among populations in the perennial grass *Festuca rubra*

Journal of
Plant Ecology

VOLUME 12, NUMBER 4,
PAGES 730–741

AUGUST 2019

doi: 10.1093/jpe/rtz010

Advanced Access published
7 March 2019

available online at
https://academic.oup.com/jpe

Temperature and precipitation, but not geographic distance, explain genetic relatedness among populations in the perennial grass *Festuca rubra*

Maria Šurinová[1,2], *Věroslava Hadincová*[2,*], *Vigdis Vandvik*[3,4] *and Zuzana Münzbergová*[1,2,*]

[1] *Department of Botany, Faculty of Science, Charles University, Prague, Czech Republic*
[2] *Institute of Botany, Academy of Sciences of the Czech Republic, Průhonice, Czech Republic*
[3] *Department of Biology, University of Bergen, Bergen, Norway*
[4] *Bjerknes Centre for Climate Research, Bergen, Norway*
*Correspondence address. Department of Botany, Faculty of Science, Charles University, Benátská 2, 128 01 Praha 2, Czech Republic. Tel: +420-22195-1639; Fax: +420-22195-1645; E-mail: zuzmun@natur.cuni.cz

Abstract

Aims
Knowledge of genetic structure of natural populations and its determinants may provide key insights into the ability of species to adapt to novel environments. In many genetic studies, the effects of climate could not be disentangled from the effects of geographic proximity. We aimed to understand the effects of temperature and moisture on genetic diversity of populations and separate these effects from the effects of geographic distance. We also wanted to explore the patterns of distribution of genetic diversity in the system and assess the degree of clonality within the populations. We also checked for possible genome size variation in the system.

Methods
We studied genetic variation within and among 12 populations of the dominant grass *Festuca rubra* distributed across a unique regional-scale climatic grid in western Norway, Europe and explored the importance of temperature, precipitation and geographic distance for the observed patterns. We also explored the distribution of genetic diversity within and among populations, identified population differentiation and estimated degree of clonality. The analyses used microsatellites as the genetic marker. The analyses were supplemented by flow cytometry of all the material.

Important Findings
All the material corresponds to hexaploid cytotype, indicating that ploidy variation does not play any role in the system. The results indicate that temperature and precipitation were better predictors of genetic relatedness of the populations than geographic distance, suggesting that temperature and precipitation may be important determinants of population differentiation. In addition, precipitation, alone and in interaction with temperature, strongly affected population genotypic diversity suggesting increased clonality towards the coldest and especially the coldest wettest climates. At the same time, individuals from the coldest and wettest climates also had the highest individual genetic diversity, suggesting that only the most heterozygous individuals survive under these harsh climates. Most of the genetic variation was distributed within populations, suggesting that most populations have sufficient genetic diversity to adapt to novel climatic conditions. The alpine populations, i.e. populations which are likely the most endangered by climate change, however, lack this potential due to the high levels of clonality as detected in our study.

Keywords: AMOVA, hexaploid, isolation by distance, isolation by environment, microsatellites, Poaceae, fuzzy clustering

Received: 9 May 2018, Revised: 13 February 2019, Accepted: 22 February 2019

© The Author(s) 2019. Published by Oxford University Press on behalf of the Institute of Botany, Chinese Academy of Sciences and the Botanical Society of China.
All rights reserved. For permissions, please email: journals.permissions@oup.com

INTRODUCTION

Understanding the ability of organisms to respond to rapid environmental change is critically important under current changing climates. Yet, we know very little about the background genetic patterns that define the potential of species to adapt to these changes (Bragg *et al.* 2015; Sexton *et al.* 2014). In the short-term, and for changes within the range of current environmental variability, plastic responses and tolerance of environmental variation are sufficient to allow population survival. However, genetic adaptation is expected to be the most important mechanism allowing species to deal with long-term climatic changes and novel climates (Franks *et al.* 2014).

Recently, we have seen an increasing number of studies exploring genetic patterns within populations distributed along climatic gradients to study the importance of climate in species adaptation (reviewed in Halbritter *et al.* 2018; Sexton *et al.* 2014). Most commonly, altitudinal gradients are used as climatic gradients assuming that the climate of higher elevations will correspond to that of lower elevations in the future, leading to upward migration of species. These studies indeed showed that populations are genetically differentiated along elevational gradients, but the degree of the differentiation strongly varies among systems (Brito *et al.* 2016; Byars *et al.* 2009; Cui *et al.* 2017; Frei *et al.* 2014; Gabel *et al.* 2017; Hamann *et al.* 2016; Sexton *et al.* 2016; Zhao *et al.* 2016).

While these studies generally demonstrate that elevation is an important driver of between-population differentiation, most of these studies are unable to detect the specific climatic factors underlying this differentiation. It is also often hard to identify to what extent the observed differentiation may be caused by simultaneous geographic differentiation. This is because different climatic variables are often correlated to each other and to geographic position of the populations and separation of the different factors is thus difficult (Still *et al.* 2005). Distant populations are likely to be more genetically differentiated from each other due to purely neutral processes such as genetic drift (Halbritter *et al.* 2018). Proper separation of climate and space is thus needed to understand the adaptive differentiation of the populations due to climate and to separate this from the neutral processes. In addition, range of previous studies demonstrated that the species response to climate is affected by simultaneous effects of temperature and precipitation (Meineri *et al.* 2013; Münzbergová *et al.* 2017; Suseela *et al.* 2012). Their separation is, however, crucial as with climate change temperature and precipitation are not chaining in concert, but rather create different novel combinations of different climatic factors (IPCC 2014).

In a recent review of studies attempting to separate the isolation by distance and the isolation by environment, Sexton *et al.* (2014) demonstrated that both types of isolation are important in structuring patterns of genetic diversity, with almost 75% of species showing some environmental effects. Despite the wealth of studies on the topic, the evidence of the effect of specific climatic variables on differentiation of plant populations is very scarce. In many studies, the effects of climate also became non-significant after accounting for geographic patterns (Still *et al.* 2005, but see Temunović *et al.* 2012).

Grasses represent a dominant group of organisms in many terrestrial habitats and thus are determinants of composition and functioning of many terrestrial communities (Gibson 2009). Understanding factors determining their differentiation is thus especially important as the changes in their performance are likely to scale up to changes in the whole ecosystems. Previous studies on various grass models indicated variable results ranging from systems showing high between-population genetic differentiation (Hamann *et al.* 2016; Rudmann-Maurer *et al.* 2007) to systems showing extensive gene flow among the grass populations (Clark *et al.* 2016; Volis *et al.* 2016). Despite this variability, many systems showed high ability of grasses to adapt to local conditions (Arnesen *et al.* 2017; Gray *et al.* 2014; Hamann *et al.* 2016; Liu *et al.* 2016; Meyer *et al.* 2016; Rudmann-Maurer *et al.* 2007). Most of these studies were, however, unable to unequivocally separate the environmental and geographic effects as these factors were intercorrelated (but see Hübner *et al.* 2009).

In a range of model systems, both grass (Manzaneda *et al.* 2012; McAllister *et al.* 2015) and non-grass (Paule *et al.* 2017; Thompson *et al.* 2014), it has been suggested that population differentiation along environmental gradients may be linked to differentiation in ploidy level. Individuals of different ploidy levels may significantly differ in levels of heterozygosity and thus in their overall genetic diversity (Münzbergová *et al.* 2013; Eliášová *et al.* 2014; Mairal *et al.* 2018) and also in a range of traits related to performance, including the degree of clonality (Baldwin and Husband 2013; Münzbergová 2007; Roberts *et al.* 2016). Understanding possible cytotype variation should thus be a key first step in studies of species adaptation and adaptive capacity along environmental gradients.

We studied genetic differentiation of populations of a dominant perennial grass, *Festuca rubra*, distributed along factorially crossed temperature and precipitation gradients in western Norway. Thanks to their factorial arrangement, these two climatic gradients are independent of each other allowing to study their independent effects. *Festuca rubra* is a species known to occur in three ploidy levels (tetraploid, hexaploid and octoploid) across its European distribution range (Dirihan *et al.* 2016). In a previous study, we demonstrated high level of genetic differentiation between these populations in a wide range of growth-related traits (Münzbergová *et al.* 2017). This variation was closely related to differences in precipitation and temperature among populations. As a next step, we now explore the cytotypic variation within our model and distribution of genetic variation within and between populations. Specifically, we ask the following questions: (i) What is the ploidy and genetic variation within and among populations? (ii) To what extent is the genetic variation related to

temperature, precipitation and their interaction and to geographic position of the populations?

Based on distribution of *F. rubra* cytotypes in Europe (Dirihan *et al.* 2016), we hypothesized that our populations will contain mostly hexaploid individuals. Based on previous studies on perennial grasses cited above, we also hypothesized that genetic diversity will be mostly found within populations with little differentiation between populations. As spatial and environmental distances are uncorrelated in our study system, we expected that it will be possible to separate the effects of geographic position and the environment on differentiation of the populations with environmental similarity being the stronger predictor of the genetic patterns. We also hypothesized that the populations from the extreme cold alpine conditions will be the most genetically differentiated from all the other populations, reflecting the patterns in species performance (Münzbergová *et al.* 2017). Finally, hypothesized that populations from the coldest environments will show lower overall genetic diversity and higher levels of clonal growth than the other populations.

MATERIALS AND METHODS

Studied system

We used *F. rubra* L. as a model species in our study. *Festuca rubra* L. is a common perennial grass species of temperate to alpine grasslands in Europe. It reproduces by seeds as well as vegetatively, producing both intravaginal and extravaginal tillers on rhizomes (Münzbergová *et al.* 2017).

The experimental plants were collected along a natural climatic grid established in western Norway (the SeedClim Grid, see Klanderud *et al.* 2015). It consists of 12 grassland localities representing three levels of summer temperature combined with each of four levels of mean annual precipitation. The sampling was set up to achieve means of the four warmest months for individual locality types of ca. 6.5°C (referred to as alpine, ALP), 8.5°C (sub-alpine, SUB) and 10.5°C (boreal, BOR) and annual precipitation of ca. 600 (1), 1300 (2), 2000 (3) and 2700 (4) mm, (Klanderud *et al.* 2015; Meineri *et al.* 2014; Table 1). The target communities are grazed meadows [Potentillo-Festucetum ovinae; G8 *sensu* (Fremstad 1997)] occurring on south-west facing (with the exception of one site (BOR 3), which was exposed to the east), shallow slopes (5–20°) with relatively base-rich bedrock. Sites were selected specifically to ensure that grazing regime and grazing history, bedrock, slope, aspect and vegetation types are as similar as possible. Geographical distance between sites is on average 15 km and ranges from 0.65 km (BOR2 and SUB2) to 175 km (BOR1 and BOR4; Meineri *et al.* 2014).

Plants from these locations were collected for our previous study (Münzbergová *et al.* 2017) in 2014, except population ALP2, which was sampled in 2016. From each of the 12 *F. rubra* natural populations, living plants were collected in the field at least 1 m apart, transplanted into pots in the experimental garden of the Institute of Botany, Czech Academy of Science, Průhonice, Czech Republic and cultivated to check their identity and assess the cytotype. We recorded exact location of each sampled plant within the population using a combination of GPS points and a measuring tape. Fresh leaves were collected from all the *F. rubra* plants established in our experimental garden (17 to 58 individuals per population, Table 1) and the leaves were dried in silica gel. The number of plants per population varies because some of the originally collected plants were found not to belong to *F. rubra* as determined after they grew in our experimental garden. This subsequent control of determination on the garden-grown plants was important as *F. rubra* grew intermixed with other *Festuca* species in the field and it was sometimes hard to distinguish

Table 1: geographic information and genetic and genotypic diversity of *Festuca rubra* populations

Abbrev.	Latitude	Longitude	Altitude (m)	Temp. level (°C)	Precip. level (mm)	Sample size	Alleles per pop.	Private alleles	Alleles per ind.	Rarified allelic richness	Genotypic diversity	Clones
ALP1	8.123	61.024	1208	6.5	600	29	33	0	12.65	6.99	0.90	2
ALP2	7.276	60.823	1097		1300	17	36	1	10.17	7.80	0.88	2
ALP3	7.170	60.831	1213		2000	32	31	1	12.00	6.07	0.59	6
ALP4	6.414	60.933	1088		2700	31	23	0	14.22	4.42	0.16	3
SUB1	8.706	60.820	815	8.5	600	40	44	2	12.95	8.13	0.90	4
SUB2	7.175	60.875	700		1300	58	46	3	11.93	7.31	0.90	3
SUB3	6.629	61.086	769		2000	27	36	0	11.85	7.22	0.67	7
SUB4	6.515	60.544	797		2700	45	39	0	11.97	7.52	0.80	7
BOR1	9.078	61.035	589	10.5	600	30	38	3	12.36	6.71	0.80	5
BOR2	7.169	60.880	474		1300	36	42	0	11.83	7.39	0.97	1
BOR3	6.337	60.665	431		2000	61	43	0	11.75	7.92	0.79	8
BOR4	5.964	60.690	346		2700	37	39	1	11.83	7.14	0.70	5

Genotypic diversity is the ratio between number of unique genotypes and ramets sampled, clones give number of genotypes with more than one sampled ramets. Temperature (means of the four warmest months) and precipitation level (mean annual rainfall) represent position of the populations along the climatic grid.

small sterile individual ramets to species under the field conditions. The identity of our sampled plants was also confirmed by flow cytometry (see below). The leaf samples and DNA extracts are deposited at the Institute of Botany, Czech Academy of Science, Průhonice, Czech Republic.

DNA extraction and standardization

Total genomic DNA was extracted from dry leaves stored in silica gel using a silica-based DNeasy Plant Mini kit (QIAGEN, Dusseldorf, Germany). Quality and quantity of extracted DNA was measured by NanoDrop 2000 (Thermo Scientific, Waltham, MA), and all the samples were normalized to 20 ng/µl for subsequent polymerase chain reaction (PCR).

Chromosome counts and flow cytometry

Somatic chromosome counts were obtained in metaphase spreads using root tips of two adult plants (one individual from SUB1 and SUB4 population) based on the methodology described in Krahulcova (2003). The low number of samples is given by the high work load required to count chromosomes in a single sample. These samples served as reference for the flow cytometry. The samples were selected to cover the extremes of genome size detected by the flow cytometry.

All individuals used for the genetic analyses were also used for the flow cytometry, i.e. 443 individuals. For flow cytometric analysis, nuclear suspensions were obtained by chopping leaf tissue of *F. rubra* with a razor blade in a Petri dish containing 0.5 ml of Otto I buffer (0.1 M citric acid, 0.5% Tween 20). *Pisum sativum* 'Ctirad' was used as the internal reference standard (2C = 8.84 pg DNA). The nuclear suspension obtained was filtered with a nylon filter and 1-ml Otto II buffer (0.4 M $Na_2HPO_4.12H_2O$) with DAPI (4 µg/m) and β-merkaptoetanol (2 µl/ml) was added to stain the DNA. All samples (443 individuals from 12 populations) were analyzed in a Partec Ploidy Analyzer II flow cytometer and the results were checked using Partec FloMax software v2.0 (Partec GmbH, Münster, Germany). At least 3000 particles were analyzed per sample. Usually, bulk samples of up to 5 plants from the same population were measured. For each sample, the DNA ploidy level was determined based on the sample/standard ratio (Castro *et al.* 2012).

Microsatellites amplification

Four microsatellite loci were amplified using *Festuca* and *Festuca–Lolium* complex-specific primers HVM20 and HVM3 (Fu *et al.* 2006) and B4-D9 and B3-B8 (Lauvergeat *et al.* 2005). While it may seem that the number of microsatellites is too low, we detected 63 alleles in these loci in total. This high number is given by the hexaploid nature of our individuals. Our preliminary tests indicated that these numbers are sufficient to reliably distinguish among clones in the species (Hadincová *et al.* in preparation). Specifically, the study of Hadincová *et al.* (in preparation) indicated that these markers can unequivocally distinguish between 20 different individuals of the species sampled within a single grassland of 100 m × 100 m and successfully used these markers to distinguish the genotypes within an experimental array established using the material.

Multiplex PCR was designed combining all microsatellite primers to one reaction. QIAGEN Multiplex PCR Master Mix (Qiagen, Hilden, Germany) containing HotStartTaq DNA polymerase, a multiplex PCR buffer at 6 µM $MgCl_2$, deoxynucleotide triphosphates and a factor MP that improves annealing and elongation were used.

Multiplex PCR contained 2.5 µl QIAGEN Multiplex PCR Master Mix, 0.25 µl of HVM 20 primers, 0.05 µl of HVM3, B3-B8, B4-D9 primers (10 µM each in initial volume) 0.7 µl H2O and 20 ng of DNA dissolved in 1-µl TE buffer. An initial denaturation step at 95°C for 15 min was followed by 40 cycles of denaturation (95°C for 30 s), annealing (60°C for 90 s) and extension (72°C for 60 s) steps, and a final extension step at 72°C for 10 min. A 1-µl aliquot of the PCR product was mixed with 11 µl of a 120: 1 solution of formamide and size standard GeneScan 500 LIZ (Life Technologies, Foster City, CA). Fragment lengths were determined by capillary gel electrophoresis with an ABI 3130 Genetic Analyzer using Gene Mapper 4.0 (Life Technologies) and the peaks were scored manually.

To check the reproducibility of our microsatellite analyses, we selected three individuals per population and repeated the microsatellite analyses of these individuals in three independent runs for each individual. The results indicated 100% match in the patterns, suggesting that our microsatellite analyses are highly reproducible. The primary data from the genetic analyses are in online supplementary material 1.

Data analysis

Because of the hexaploid nature of our data, we were not able to score the exact proportions of each allele in the heterozygotes. As a result, the alleles were only coded as present or absent. To account for unequal number of samples per population, genetic diversity of each population was expressed as rarefied allelic richness (Ar) assessed using SPAGeDi (Hardy, Vekemans 2002). We also calculated average number of alleles per individual as an additional measure of genetic diversity of the populations.

Individuals sharing multi-locus genotypes (MLG) were considered to be clones as we previously proved that our marker is suitable to do this for the species (Hadincová et al. in preparation). Number of unique genotypes divided by number of samples analyzed per population was used to express genotypic diversity of each population. As our model is a hexaploid and it was not possible to score allele frequencies, we could not use the standard methods to calculate the probabilities of each MLG arising through sexual reproduction (Arnaud-Haond and Belkhir 2007; Parks and Werth 1993). Because our MGLs were characterized by range of differences across different loci, we expect that it is unlikely that segregation and recombination in the course of sexual reproduction

of the species would give rise to identical heterozygous MLGs (Tibayrenc *et al.* 1990).

Total gene diversity (H_T) was estimated as the expected heterozygosity (H_E) value for the whole data set, and the mean genetic diversity within populations (H_S) was calculated as the mean of the mean H_E value of all the populations in the data set. Genetic diversity among populations (D_{ST}) was calculated as $D_{ST} = H_T - H_S$. Because our plants are hexaploid leading to unknown dosage of single alleles, the H_E values are values of expected proportion of individuals bearing more than one allele and thus correspond to zygotic heterozygosity (Münzbergová *et al.* 2013). Such a measure of expected heterozygosity can be applied to any ploidy level (Bayer 1989; Nei 1973).

To evaluate distribution of genetic variance among and within populations, analysis of molecular variance (AMOVA) was performed using GenAlEx 6 (Peakall, Smouse 2012). Significance of the AMOVA results was tested using 999 random permutations of the data. In addition to analyzing the whole data set, we also calculated H_T, H_S and D_{ST} and performed AMOVA for populations within each temperature and within each precipitation level separately. In this way, we could compare the level of between-population differentiation within each of those levels.

To estimate population structure, we used fuzzy *c*-means clustering analysis. We used this analysis instead of the commonly used STRUCTURE 2.3.4. software (Pritchard *et al.* 2000) as STRUCTURE is not designed for analyses of hexaploid species. We, however, include the STRUCTURE analyses for comparison and show its results including detailed methodology in online supplementary material 2. Fuzzy *c*-means clustering is the extension of the *k*-means algorithm and is working without any biological assumptions (Bezdek 1981) and has been previously used to analyze microsatellite and AFLP data (Arrigo *et al.* 2010, 2011; Burnier *et al.* 2009). The procedure has been in detailed described in Arrigo *et al.* (2011). In brief, the algorithm follows an iterative process to allocate individuals within a predefined number of groups (K) in a way to minimize the intra-group variance. The computation relies on a fuzzification parameter (*r*, ranging between one and infinity). R = 1 produces non-fuzzy clusters, while larger *r* values increase the fuzziness of clustering until all individuals are equally assigned to all the groups. As the choice of *r* is partly subjective, we tested a range of values between 1.05 and 2 to identify the highest value that allows assignment of the individuals into the groups (at least 10% difference of assignment probability to each of the groups for at least 90% of the individuals). Based on this criterion, we selected $r = 1.1$ for the final analyses. Maximum iterations were set to 10 000, as further increase of the value did not change the results. The analyses were done using cmeans function in e1071 library in R, using ufcl method. We identified the optimal number of clusters (K) by testing values of K 2–12 and comparing the models using partition coefficient.

To estimate the proportion of genetic differentiation among populations explained by geographic and environmental distance, we constructed several dissimilarity matrices. A pairwise genetic distance matrix between all samples was calculated using Bruvo distance (Bruvo *et al.* 2004), which takes distances between microsatellite alleles into account without the knowledge of allele copy number using R package Polysat, (Clark and Jasieniuk 2011). Polysat is a software for analyzing polyploid microsatellite data. It handles genotype data of any ploidy and assumes that the allele copy number is always ambiguous in partial heterozygotes. Geographic and environmental distances were based on Euclidean distance. For environmental distance, we separately created matrix for difference in temperature and in precipitation and also a matrix combining both of the variables. We used Mantel test with 999 permutations to do the analyses (R package vegan). First, we tested the effect of geographic and each separate environmental matrix on the genetic matrix. Then we used partial Mantel test and tested the effect of each environmental matrix on the genetic structure after accounting for the geographic structure, and the effect of geographic structure on the genetic structure after accounting for the environmental structure. We also performed partial tests of each of the environmental variables after accounting for all the other environmental variables. As these tests yielded little additional information to the tests presented above, they are not described further. In addition to the analyses across all the populations, we also tested the relationship between genetic and geographic distance within each population separately, to explore if there was any internal genetic structure within the populations. Environmental data were not tested at this scale as they were identical within populations. The use of Mantel test for such analyses has been recently widely discussed (Balkenhol *et al.* 2009; Graves *et al.* 2013; Guillot and Rousset 2013) suggesting possible alternative approaches (Bradburd *et al.* 2013). However, these alternative approaches require high number of parameters and have not been sufficiently tested for their predictive power. We thus follow the suggestion of Vergara *et al.* (2017) that Mantel test is still the best, most robust and easily comparable approach to assess the importance of isolation by distance and by environment.

Finally, we tested the relationship between temperature and precipitations of population origin and their interaction on rarified allelic richness, mean number of alleles per individual and genotypic diversity per population using ANOVA. All the dependent variables followed normal distribution and showed homogeneity of variance justifying use of ANOVA for this purpose.

RESULTS

Chromosome counts and ploidy level in *F. rubra*

Flow cytometry using all the 443 *F. rubra* individuals resulted in high-resolution histograms with low coefficient of variation and negligible background noise. There was very

little variation in the relative genome size values between the samples (mean ± SE, 1.1161 ± 0.00088, N = 443). The two samples used for karyological analyses were selected as the extreme values of the relative genome size (1.082 and 1.188). The same chromosome number ($2n$ = 42) was identified in both samples. This corresponds to hexaploid cytotype.

Genetic and genotypic diversity

A total of 63 alleles at four microsatellite loci were detected across the 443 studied individuals. The number of alleles within each of the 12 populations ranged between 23 (ALP4) and 46 (SUB2). A total of 11 private alleles (i.e. alleles occurring just in a single population) distributed across six populations were detected (Table 1). Rarefied allelic richness was the lowest in population ALP4 and the highest in SUB1. The ALP4 population had the highest number of alleles per individual, while ALP3 population had the lowest number of alleles per individual. ALP4 population also had the lowest genotypic diversity, while BOR2 had the highest genotypic diversity (Table 1).

While most ramets sampled represented unique genotypes, in all populations we found some ramets sharing the same genotype, likely representing clones. Number of genotypes with multiple sampled ramets ranged from 1 in BOR2 to 8 in BOR3 (Table 1). While in most cases only two ramets had the same MLG, this number raised to 16 in one occasion in ALP4 population.

Genotypic diversity, but not rarified allelic richness and number of alleles per individual, was significantly negatively related to precipitation. There was no main effect of temperature on any of the three variables, but there was a significant interactive effect of temperature and precipitation on genotypic diversity and rarified allelic richness (Table 2) with the negative effect of precipitation being the strongest in the coldest environments.

Population structure and differentiation

AMOVA analyses of the total data set showed that the majority of the genetic variation was within populations (Table 3). Similar results were detected in separate analyses of single precipitation and temperature levels (Table 3). The highest among-population differentiation accompanied with the highest D_{ST} value was found among the populations in the coldest temperature zone and in the wettest precipitation zone (Table 3).

Fuzzy clustering analysis identified K = 7 as the optimal number of clusters. The first cluster represented majority of individuals belonging to ALP4 population. The second cluster represented majority of individuals belonging to ALP3 population. The other clusters contained individuals from most populations (Fig. 1). This result corresponds to results of the STRUCTURE analysis identifying K = 6 as the optimal number of clusters and very similar distribution of individuals among the clusters (online supplementary material 3). To check the effects of clonality on the results of this analysis, we linked the data on cluster identity to data on clonality. While both of the deviating populations according to the cluster analyses (ALP3 and ALP4) are partly clonal (especially ALP4), the existence of the distinct clusters is not caused primarily by clonality as individuals belonging to different clones fall into the same dominant cluster. Also, some individuals from other populations were identified as part of these clusters (online supplementary material 2).

There was no significant relationship between genetic and geographic distance matrices for the total data set (Table 4). Genetic distances were, however, significantly positively related to all the three environmental distances with the distance combining temperature and precipitation having the strongest relationship, and precipitation distance having the weakest relationship. The results of all the partial tests were largely similar to the tests without the covariates (Table 4). There was also a significant positive correlation between geographic and genetic distances within ALP3, SUB4, BOR3 and BOR4 populations but not within the other populations. The strongest increase of genetic distance with geographic distance was detected for the ALP3 population (online supplementary material 3).

DISCUSSION

All the studied individuals were hexaploid with 42 chromosomes. This corresponds to the study of Dirihan *et al.* (2016), which found the hexaploid cytotype to be the most common in *F. rubra* in Europe and prevailing also in northern Europe (Loureiro *et al.* 2007; Sampoux and Huyghe 2009). These findings suggest that possible differentiation of the plants within and between populations is not linked to cytotype changes. An opposite pattern, i.e. between population differentiation in the ploidy level at a landscape scale, was found in other

Table 2: effect of temperature, precipitation and their interaction on different population-level genetic characteristics assessed using ANOVA

	Alleles per ind.		Rarified allelic richness		Genotypic diversity	
	F	*P*	*F*	*P*	*F*	*P*
Temperature	0.21	0.660	3.46	0.100	3.62	0.094
Precipitation	0.06	0.814	2.79	0.134	**11.17**	**0.010**
Temp. × Prec.	1.74	0.224	**5.82**	**0.042**	**5.56**	**0.046**

Significant values ($P \leq 0.05$) are in bold.

Table 3: summary statistics describing genetic differentiation among populations within the total data set and within the temperature and the precipitation levels

	Data set	H_T	H_S	D_{ST}	Within pop.	Among pop.
	Total	0.8227	0.7906	0.0321	82	18
Temperature	Alpine	0.8084	0.7675	0.0409	62	38
	Sub-alpine	0.8338	0.8246	0.0092	90	10
	Boreal	0.8245	0.8064	0.0181	92	8
Precipitation	2700	0.8357	0.8006	0.0351	73	27
	2000	0.8330	0.8084	0.0246	89	11
	1300	0.7986	0.7930	0.0056	95	5
	600	0.8277	0.8115	0.0162	92	8

D_{ST} represents genetic diversity among the populations, H_S is mean genetic diversity within the populations and H_T is total gene diversity within the category. The last two columns represent within- and among-population variation assessed using AMOVA; all these values are significantly deviating from zero ($P \leq 0.05$).

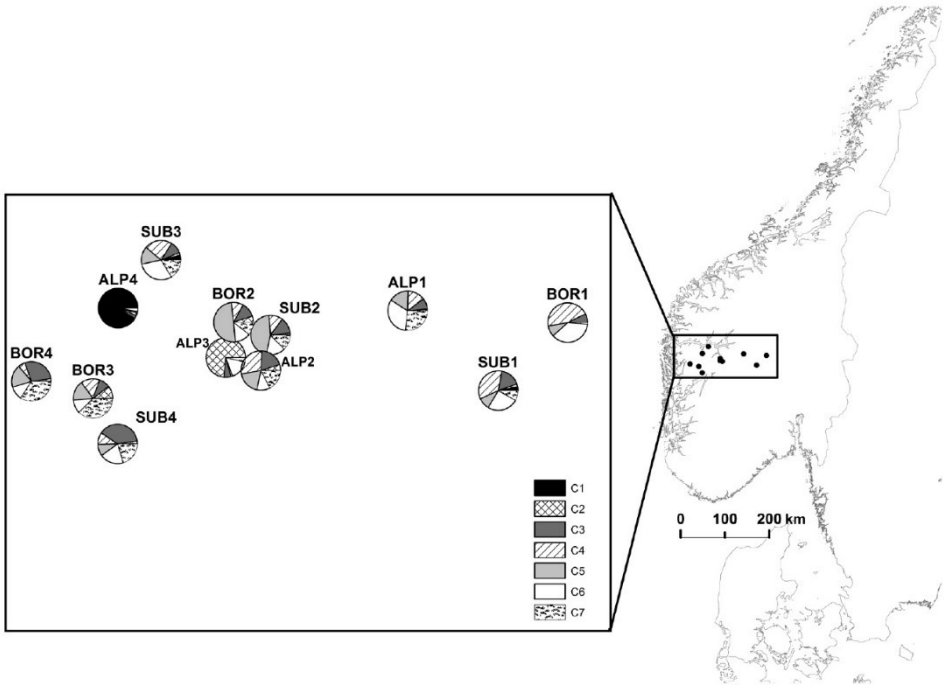


Figure 1: results of fuzzy clustering analysis representing distribution of individuals belonging to different clusters (C1–C7) within the studied populations. Similar figure produced by STRUCTURE analysis can be found in online supplementary material 2, Fig. A (See online supplementary material for a colour version of this figure.).

Festuca species (Ceccarelli *et al.* 1992; Smarda and Bures 2006) as well as in a range of other species of different families (Castro *et al.* 2012; Paule *et al.* 2017). While cytotype diversity contributed to high genetic diversity in a range of systems (Eliášová *et al.* 2014; Luttikhuizen *et al.* 2007; Münzbergová *et al.* 2013; Rosche *et al.* 2016), it was clearly not the case in the system studied here.

Genetic diversity detected in our model populations roughly corresponded to genetic diversity in a range of previous studies on various grass species (Hamann *et al.* 2016; Rudmann-Maurer *et al.* 2007). Genetic as well as genotypic diversity was the lowest, while number of alleles per individual was the highest, in the population from the most extreme wet cold environment (ALP4). Locality of this population is characterized

Table 4: the correlation between genetic distance and geographic (Geo.) or environmental distance (based on temperature – Temp., precipitation – Prec. or both Temp. × Prec. defined under second matrix) after accounting for the variable defined under Covariate assessed using Mantel or partial Mantel test

Second matrix	Covariate	*r*	*P*
Geo.	—	0.010	0.255
Temp.	—	**0.112**	**0.001**
Prec.	—	**0.085**	**0.001**
Temp. × Prec.	—	**0.143**	**0.001**
Temp.	Geo.	**0.111**	**0.001**
Prec.	Geo.	**0.085**	**0.001**
Temp. × Prec.	Geo.	**0.143**	**0.001**
Geo.	Temp.	0.006	0.336
Geo.	Prec.	0.002	0.393
Geo.	Temp. × Prec.	−0.005	0.471

Significant correlations ($P \leq 0.05$) are shown in bold.

by the shortest growing season (Delnevo *et al.* 2018). Similarly, low genetic diversity in the most extreme alpine environment was demonstrated by Brito *et al.* (2016), in *Tibouchina pulchra* (Melastomataceae) growing along an elevation gradient in montane Atlantic rainforest. As they did not find any differences in the genetic diversity between pollen arriving to the different sites, their results suggest that the lower diversity in high elevation was not due to the lack of gene flow, but rather due to the strong selection pressure of the environment. Alternatively, lower genetic diversity in the alpine populations may also be explained by the fact that the alpine populations represent marginal habitats (McMinn *et al.* 2017). Lower genetic diversity at the marginal habitats may be attributed to a range of processes including founder effects, strong genetic drift due to smaller population size, limited gene flow as well as strong selection under the most extreme environments.

High number of alleles per individual, reflecting high levels of heterozygosity, in the extreme ALP4 population is in line with the results of Luijten *et al.* (2000), demonstrating that only the most heterozygous individuals were able to survive in the most extreme conditions (in their case the smallest populations). This is likely a consequence of the general trend that more heterozygous individuals have higher fitness (reviewed in Reed and Frankham 2003, later e.g. Dostálek *et al.* 2010, but see Rodríguez-Quilón *et al.* 2015). The strength of the relationship between heterozygosity and fitness may, however, vary with environmental stress (Ferrer *et al.* 2016; Vranckx *et al.* 2014).

Genotypic diversity decreased with increasing precipitation and the decrease was stronger in the colder regions. Low genotypic diversity likely reflects extensive clonal growth in the species. Increased clonal growth in the wet environments and especially in the coldest and wettest population is in line with the fact that the wet environment represents stressful conditions for the species. Plants at this location also produced the highest amounts of rhizomes and extravaginal ramets (Münzbergová *et al.* 2017). Note that we consider the wettest environment as stressful as our wettest sites have 2700 mm of annual rainfall and are thus very wet, while the dry sites have 600-mm annual rainfall and correspond to what most studies view as mesic or wet. As a high fraction of the precipitation occurs in winter, the wet sites have extensive snow cover leading to strongly reduced duration of the vegetation season especially in the colder part of the gradient (Delnevo *et al.* 2018). In addition, the plants at the wettest sites may also suffer from stress due to nutrient leaching and low light levels caused by frequent cloud cover (www.met.no). The observed pattern is thus in line with the fact that clonality is an efficient mechanism allowing species to cope with various environmental stresses (Brezina *et al.* 2006; Prati and Schmid 2000; Santamaria 2002; Vandepitte *et al.* 2010). Indeed, Wookey *et al.* (1995) demonstrated that reducing environmental stress leads to reduced clonal reproduction and increased generative reproduction in arctic environments. These results support our findings of increased clonality in alpine environments. This may indicate that climate warming may increase the importance of generative reproduction in comparison to clonality. As generative reproduction is crucial for generating new genotypes, this may increase the ability of the species to adapt to the novel climates.

The absence of the simple linear relationship between the population genetic diversity and temperature and precipitation is also in line with other studies demonstrating that such relationships may be strongly non-linear (Byars *et al.* 2009) or that such effects may be weaker than effects of local habitat conditions (Stocklin *et al.* 2009). It, however, contrasts to results of our previous studies demonstrating that temperature and precipitation of the habitat of origin as well as their interactions were significant predictors of performance of the same model species from the same system in controlled conditions (Münzbergová *et al.* 2017; Münzbergová and Hadincová 2017) and also to the patterns of genetic similarity among populations studied here (discussed below). This seemingly contradictory finding is likely linked to the fact that the phenotypic traits as well as specific genetic profiles are more likely under selection pressure than genetic diversity.

AMOVA analyses and D_{st}/H_T comparison of the total data set showed that the variation within populations was significantly higher than the variation among populations. This agrees with other studies in grasses (McGrath *et al.* 2007; Rudmann-Maurer *et al.* 2007) and indicates low among population variation. High ploidy level, wind-pollination, long-distance dispersal ability and long life span may contribute significantly to the low genetic differentiation among the populations in the system (Andreakis *et al.* 2009; Tanhuanpää and Manninen 2012; Thiel-Egenter *et al.* 2009).

Comparison of the populations within the single temperature and precipitation levels demonstrated that the highest differentiation can be found in the most extreme populations,

i.e. populations within the coldest and within the wettest zone. This is likely linked to the strongly differentiated ALP4 (and partly also ALP3) population visible also in the fuzzy clustering analysis. It indicates that under stress, the plants are more likely to experience strong selection pressure leading to strong between population differentiation (Eckert *et al.* 2008). This pattern in ALP4 is unlikely to be caused by the high clonality in this population as the dominant cluster includes individuals of different clones.

High level of genetic diversity within populations is also reflected in the absence of any isolation by distance for the total data set. This result contrasts to a range of previous studies (as reviewed in Sexton *et al.* 2014), suggesting that isolation by distance is a common pattern in natural populations. This is likely because our study area shows extreme climatic and topographic variation. In addition, due to the varied topography, geographic distance may in fact not well reflect the connectivity among the populations. The former explanation is supported by significant relationship between genetic and all the environmental distances with the patterns almost unchanged after accounting for the geographic distance. The review of Sexton *et al.* (2014) in fact also indicated that environmental distances were good predictors of genetic differentiation in many cases.

Neutral genetic markers such as microsatellites may miss important functional differentiation among populations (Bekessy *et al.* 2003) and more direct measures of functional relationships among populations may be more useful to study the effects of climate (Bekessy *et al.* 2003; Yardeni *et al.* 2016). Our data, however, suggest that the effects of climate on genetic differentiation of natural populations may be detected even using microsatellites, i.e. a presumably neutral genetic marker. In line with this, Stojanova *et al.* (2018) demonstrated direct link between specific microsatellite markers and functional traits in this system. In addition, Münzbergová *et al.* (2019) demonstrated that the genetic relatedness based on these markers also well explains epigenetic differentiation of the plants.

The predictive power of our data may be limited by use of only four microsatellite markers. While it may seem that the number of microsatellites is too low, we detected 63 different alleles in these loci in total. This high number is given by the hexaploid nature of our individuals. Our preliminary tests indicated that this number is sufficient to reliably distinguish individual clones in our system (Hadincová et al. in prep). Based on this, we suggest that our data were variable enough to make reliable inferences about the genetic relatedness of populations as well as to distinguish clones.

CONCLUSION

Our study demonstrated that despite most genetic diversity being distributed within populations, the genetic differentiation of the populations is significantly linked to environment of origin. This suggests that both temperature as well as precipitation represent likely determinants of genetic relatedness in the system and of the level of clonal growth of the species. The populations of the species thus seem to have adapted to climatic conditions in the past. However, the current climate change is operating at a much faster rate and it is thus unclear if the populations will be sufficiently quick to adapt to these novel conditions. The high levels of within-population genetic diversity together with the fact that genetic variation is linked to the phenotypic and epigenetic variation as shown in our previous studies suggest that the populations should be able to adapt to the novel climate. However, the high levels of clonality in the most extreme populations reduce the effective population size and increase the generation times reducing the expected adaptive potential of the species. Future studies thus need to directly assess the efficiency of selection under different climates to uncover the real potential of the species to cope with the changing climates.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of Plant Ecology* online.

FUNDING

The study was supported by project Grant Agency of the Czech Republic (GAČR) 19-00522S and partly by institutional research projects RVO 67985939 and Ministry of Education of the Czech Republic (MŠMT). The SEEDCLIM Climate Grid field sites in western Norway and the climate and environmental data are funded by the Norwegian Research Council projects NORKLIMA 184912 and KLIMAFORSK 244525 (PI V. Vandvik).

ACKNOWLEDGEMENTS

We thank V. Jarolímová for counting the chromosomes, Z. Líblová for the flow cytometric analyses, J. Knappová for help with collecting the plants, I. Chmelařová and I. Jarošincová for help with plant cultivation, the SEEDCLIM project team for assistance with collecting climatic and environmental data, and the members of the PopEkol discussion group and two anonymous reviewers for useful comments on the previous version of the manuscript.
Conflict of interest statement. None declared.

REFERENCES

Andreakis N, Kooistra WH, Procaccini G (2009) High genetic diversity and connectivity in the polyploid invasive seaweed *Asparagopsis taxiformis* (Bonnemaisoniales) in the Mediterranean, explored with microsatellite alleles and multilocus genotypes. *Mol Ecol* **18**:212–26.

Arnaud-Haond S, Belkhir K (2007) GENCLONE: a computer program to analyse genotypic data, test for clonality and describe spatial clonal organization. *Mol Ecol Notes* **7**:15–7.

Arnesen S, Coleman CE, Meyer SE (2017) Population genetic structure of *Bromus tectorum* in the mountains of western North America. *Am J Bot* **104**:879–90.

Arrigo N, Felber F, Parisod C, *et al.* (2010) Origin and expansion of the allotetraploid *Aegilops geniculata*, a wild relative of wheat. *New Phytol* **187**:1170–80.

Arrigo N, Guadagnuolo R, Lappe S, *et al.* (2011) Gene flow between wheat and wild relatives: empirical evidence from *Aegilops geniculata*, *Ae. neglecta* and *Ae. triuncialis*. *Evol Appl* **4**:685–95.

Baldwin SJ, Husband BC (2013) The association between polyploidy and clonal reproduction in diploid and tetraploid *Chamerion angustifolium*. *Mol Ecol* **22**:1806–19.

Balkenhol N, Waits LP, Dezzani RJ (2009) Statistical approaches in landscape genetics: an evaluation of methods for linking landscape and genetic data. *Ecography* **32**:818–30.

Bayer RJ (1989) Patterns of isozyme variation in the *Antennaria rosea* (Asteraceae, Inuleae) polyploid agamic complex. *Syst Bot* **14**:389–97.

Bekessy SA, Ennos RA, Burgman MA, *et al.* (2003) Neutral DNA markers fail to detect genetic divergence in an ecologically important trait. *Biol Cons* **110**:267–75.

Bezdek J (1981) *Pattern Recognition With Fuzzy Objective Function Algorithms*. New York: Plenum Press.

Bradburd GS, Ralph PL, Coop GM (2013) Disentangling the effects of geographic and ecological isolation on genetic differentiation. *Evolution* **67**:3258–73.

Bragg JG, Supple MA, Andrew RL, *et al.* (2015) Genomic variation across landscapes: insights and applications. *New Phytol* **207**:953–67.

Brezina S, Koubek T, Münzbergová Z, *et al.* (2006) Ecological benefits of integration of *Calamagrostis epigejos* ramets under field conditions. *Flora* **201**:461–7.

Brito VLG, Mori GM, Vigna BBZ, *et al.* (2016) Genetic structure and diversity of populations of polyploid *Tibouchina pulchra* Cogn. (Melastomataceae) under different environmental conditions in extremes of an elevational gradient. *Tree Genet Genom* **12**:12.

Bruvo R, Michiels NK, D'Souza TG, *et al.* (2004) A simple method for the calculation of microsatellite genotype distances irrespective of ploidy level. *Mol Ecol* **13**:2101–6.

Burnier J, Buerki S, Arrigo N, *et al.* (2009) Genetic structure and evolution of alpine polyploid complexes: *Ranunculus kuepferi* (Ranunculaceae) as a case study. *Mol Ecol* **18**:3730–44.

Byars SG, Parsons Y, Hoffmann AA (2009) Effect of altitude on the genetic structure of an alpine grass, *Poa hiemata*. *Ann Bot* **103**:885–99.

Castro S, Loureiro J, Procházka T, *et al.* (2012) Cytotype distribution at a diploid-hexaploid contact zone in *Aster amellus* (Asteraceae). *Ann Bot* **110**:1047–55.

Ceccarelli M, Falistocco E, Cionini PG (1992) Variation of genome size and organization within hexaploid *Festuca arundinacea*. *Theor Appl Genet* **83**:273–8.

Clark LV, Dzyubenko E, Dzyubenko N, *et al.* (2016) Ecological characteristics and in situ genetic associations for yield-component traits of wild *Miscanthus* from eastern Russia. *Ann Bot* **118**:941–55.

Clark LV, Jasieniuk M (2011) POLYSAT: an R package for polyploid microsatellite analysis. *Mol Ecol Resour* **11**:562–6.

Cui D, Tang C, Li J, *et al.* (2017) Genetic structure and isolation by altitude in rice landraces of Yunnan, China revealed by nucleotide and microsatellite marker polymorphisms. *PLOS ONE* **12**:e0175731.

Delnevo N, Petraglia A, Carbognani M, *et al.* (2018) Plastic and genetic responses to shifts in snowmelt time affects the reproductive phenology and growth of *Ranunculus acris*. *Persp Plant Ecol Evol Syst* **30**:62–70.

Dirihan S, Helander M, Väre H, *et al.* (2016) Geographic variation in *Festuca rubra* L. Ploidy levels and systemic fungal endophyte frequencies. *PLOS ONE* **11**:e0166264.

Dostálek T, Münzbergová Z, Plačková I (2010) Genetic diversity and its effect on fitness in an endangered plant species, *Dracocephalum austriacum* L. *Cons Gen* **11**:773–83.

Eckert CG, Samis KE, Lougheed SC (2008) Genetic variation across species' geographical ranges: the central-marginal hypothesis and beyond. *Mol Ecol* **17**:1170–88.

Eliášová A, Trávníček P, Mandák B, *et al.* (2014) 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**:159–70.

Ferrer ES, García-Navas V, Sanz JJ, *et al.* (2016) The strength of the association between heterozygosity and probability of interannual local recruitment increases with environmental harshness in blue tits. *Ecol Evol* **6**:8857–69.

Franks SJ, Weber JJ, Aitken SN (2014) Evolutionary and plastic responses to climate change in terrestrial plant populations. *Evol Appl* **7**:123–39.

Frei ER, Hahn T, Ghazoul J, *et al.* (2014) Divergent selection in low and high elevation populations of a perennial herb in the Swiss Alps. *Alp Bot* **124**:131–42.

Fremstad E (1997) Vegetasjonstyper i Norge. *NINA Temahefte* **12**:1–279.

Fu YB, Qiu J, Peterson GW, *et al.* (2006) Characterization of microsatellite markers for rough fescue species (*Festuca* spp.). *Mol Ecol Notes* **6**:894–6.

Gabel AR, Sattler J, Reisch C (2017) Genetic variation and performance of the alpine plant species *Dianthus callizonus* differ in two elevational zones of the Carpathians. *Alp Bot* **127**:65–74.

Gibson DJ (2009) *Grasses and Grassland Ecology*. United Kingdom: Oxford University Press.

Graves TA, Beier P, Royle JA (2013) Current approaches using genetic distances produce poor estimates of landscape resistance to interindividual dispersal. *Mol Ecol* **22**:3888–903.

Gray MM, St Amand P, Bello NM, *et al.* (2014) Ecotypes of an ecologically dominant prairie grass (*Andropogon gerardii*) exhibit genetic divergence across the U.S. Midwest grasslands' environmental gradient. *Mol Ecol* **23**:6011–28.

Guillot G, Rousset F (2013) Dismantling the Mantel tests. *Methods Ecol Evol* **4**:336–44.

Halbritter AH, Fior S, Keller I, *et al.* (2018) Trait differentiation and adaptation of plants along elevation gradients. *J Evol Biol* **31**:784–800.

Hamann E, Kesselring H, Armbruster GFJ, *et al.* (2016) Evidence of local adaptation to fine- and coarse-grained environmental variability in *Poa alpina* in the Swiss Alps. *J Ecol* **104**:1627–37.

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**:618–20.

Hübner S, Höffken M, Oren E, *et al.* (2009) Strong correlation of wild barley (*Hordeum spontaneum*) population structure with temperature and precipitation variation. *Mol Ecol* **18**:1523–36.

IPCC (2014) *Climate Change 2014: Synthesis Report*. http://www.ipcc.ch/report/ar5/syr/. (15 September 2016, date last accessed).

Klanderud K, Vandvik V, Goldberg D (2015) The importance of biotic vs. abiotic drivers of local plant community composition along regional bioclimatic gradients. *PLOS ONE* **10**:e0130205.

Krahulcova A (2003) Chromosome numbers in selected monocotyledons (Czech Republic, Hungary, and Slovakia). *Preslia* **75**:97–113.

Lauvergeat V, Barre P, Bonnet M, *et al.* (2005) Sixty simple sequence repeat markers for use in the *Festuca-Lolium* complex of grasses. *Mol Ecol Notes* **5**:401–5.

Liu WS, Zhao Y, You JL, *et al.* (2016) Morphological and genetic variation along a north-to-south transect in *Stipa purpurea*, a dominant grass on the qinghai-tibetan plateau: implications for response to climate change. *PLOS ONE* **11**:e0161972.

Loureiro J, Kopecky D, Castro S, *et al.* (2007) Flow cytometric and cytogenetic analyses of Iberian Peninsula *Festuca* spp. *Plant Syst Evol* **269**:89–105.

Luijten SH, Dierick A, Oostermeijer JGB, *et al.* (2000) Population size, genetic variation, and reproductive success in a rapidly declining, self-incompatible perennial (*Arnica montana*) in The Netherlands. *Cons Biol* **14**:1776–87.

Luttikhuizen PC, Stift M, Kuperus P, *et al.* (2007) Genetic diversity in diploid vs. tetraploid *Rorippa amphibia* (Brassicaceae). *Mol Ecol* **16**:3544–53.

Mairal M, Šurinová M, Castro S, *et al.* (2018) Unmasking cryptic biodiversity in polyploids: origin and diversification of *Aster amellus* aggregate. *Ann Bot* **122**:1047–59.

Manzaneda AJ, Rey PJ, Bastida JM, *et al.* (2012) Environmental aridity is associated with cytotype segregation and polyploidy occurrence in *Brachypodium distachyon* (Poaceae). *New Phytol* **193**:797–805.

McAllister C, Blaine R, Kron P, *et al.* (2015) Environmental correlates of cytotype distribution in *Andropogon gerardii* (Poaceae). *Am J Bot* **102**:92–102.

McGrath S, Hodkinson TR, Barth S (2007) Extremely high cytoplasmic diversity in natural and breeding populations of *Lolium* (Poaceae). *Heredity* **99**:531–44.

McMinn RL, Russell FL, Beck JB (2017) Demographic structure and genetic variability throughout the distribution of Platte thistle (*Cirsium canescens* Asteraceae). *J Biog* **44**:375–85.

Meineri E, Skarpaas O, Spindelbock J, *et al.* (2014) Direct and size-dependent effects of climate on flowering performance in alpine and lowland herbaceous species. *J Veg Sci* **25**:275–86.

Meineri E, Spindelbock J, Vandvik V (2013) Seedling emergence responds to both seed source and recruitment site climates: a climate change experiment combining transplant and gradient approaches. *Plant Ecol* **214**:607–19.

Meyer SE, Leger EA, Eldon DR, *et al.* (2016) Strong genetic differentiation in the invasive annual grass *Bromus tectorum* across the Mojave-Great Basin ecological transition zone. *Biol Invasions* **18**:1611–28.

Münzbergová Z, Surinová M, Castro S (2013) Absence of gene flow between diploids and hexaploids of *Aster amellus* at multiple spatial scales. *Heredity* **110**:123–30.

Münzbergová Z (2007) Population dynamics of diploid and hexaploid populations of a perennial herb. *Ann Bot* **100**:1259–70.

Münzbergová Z, Hadincová V (2017) Transgenerational plasticity as an important mechanism affecting response of clonal species to changing climate. *Ecol Evol* **7**:5236–47.

Münzbergová Z, Hadincová V, Skálová H, *et al.* (2017) Genetic differentiation and plasticity interact along temperature and precipitation gradients to determine plant performance under climate change. *J Ecol* **105**:1358–73.

Münzbergová Z, Latzel V, Šurinová M, *et al.* (2019) DNA methylation as a possible mechanism affecting ability of natural populations to adapt to changing climate. *Oikos* **128**:124–34.

Nei M (1973) Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci USA* **70**:3321–3.

Parks JC, Werth CR (1993) A study of spatial features of clones in a population of bracken fern, *Pteridium aquilinum* (Dennstaedtiaceae). *Am J Bot* **80**:537–44.

Paule J, Wagner ND, Weising K, *et al.* (2017) Ecological range shift in the polyploid members of the South American genus *Fosterella* (Bromeliaceae). *Ann Bot* **120**:233–43.

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research–an update. *Bioinformatics* **28**:2537–9.

Prati D, Schmid B (2000) Genetic differentiation of life-history traits within populations of the clonal plant *Ranunculus reptans*. *Oikos* **90**:442–56.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**:945–59.

Reed DH, Frankham R (2003) Correlation between fitness and genetic diversity. *Cons Biol* **17**:230–7.

Roberts DG, Forrest CN, Denham AJ, *et al.* (2016) Varying levels of clonality and ploidy create barriers to gene flow and challenges for conservation of an Australian arid-zone ecosystem engineer, *Acacia loderi*. *Biol J Lin Soc* **118**:330–43.

Rodríguez-Quilón I, Santos-del-Blanco L, Grivet D, *et al.* (2015) Local effects drive heterozygosity-fitness correlations in an outcrossing long-lived tree. *Proc Biol Sci* **282**:20152230.

Rosche C, Durka W, Hensen I, *et al.* (2016) 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. *Biol Invasions* **18**:1895–10.

Rudmann-Maurer K, Weyand A, Fischer M, *et al.* (2007) Microsatellite diversity of the agriculturally important alpine grass *Poa alpina* in relation to land use and natural environment. *Ann Bot* **100**:1249–58.

Sampoux JP, Huyghe C (2009) Contribution of ploidy-level variation and adaptive trait diversity to the environmental distribution of taxa in the 'fine-leaved fescue' lineage (genus *Festuca* subg. *Festuca*). *J Biog* **36**:1978–93.

Santamaria L (2002) Why are most aquatic plants widely distributed? Dispersal, clonal growth and small-scale heterogeneity in a stressful environment. *Acta Oec* **23**:137–54.

Sexton JP, Hangartner SB, Hoffmann AA (2014) Genetic isolation by environment or distance: which pattern of gene flow is most common? *Evolution* **68**:1–15.

Sexton JP, Hufford MB, Bateman AC, *et al.* (2016) Climate structures genetic variation across a species' elevation range: a test of range limits hypotheses. *Mol Ecol* **25**:911–28.

Smarda P, Bures P (2006) Intraspecific DNA content variability in *Festuca pallens* on different geographical scales and ploidy levels. *Ann Bot* **98**:665–78.

Still DW, Kim DH, Aoyama N (2005) Genetic variation in *Echinacea angustifolia* along a climatic gradient. *Ann Bot* **96**:467–77.

Stocklin J, Kuss P, Pluess AR (2009) Genetic diversity, phenotypic variation and local adaptation in the alpine landscape: case studies with alpine plant species. *Bot Helv* **119**:125–33.

Stojanova B, Šurinová M, Klápště J, *et al.* (2018) Adaptive differentiation of *Festuca rubra* along a climate gradient revealed by molecular markers and quantitative traits. *PLOS ONE* **13**:e0194670.

Suseela V, Conant RT, Wallenstein MD, *et al.* (2012) Effects of soil moisture on the temperature sensitivity of heterotrophic respiration vary seasonally in an old-field climate change experiment. *Global Change Biol* **18**:336–48.

Tanhuanpää P, Manninen O (2012) High SSR diversity but little differentiation between accessions of Nordic timothy (*Phleum pratense* L.). *Hereditas* **149**:114–27.

Temunović M, Franjić J, Satovic Z, *et al.* (2012) Environmental heterogeneity explains the genetic structure of Continental and Mediterranean populations of *Fraxinus angustifolia* Vahl. *PLOS ONE* **7**:e42764.

Thiel-Egenter C, Gugerli F, Alvarez N, *et al.* (2009) Effects of species traits on the genetic diversity of high-mountain plants: a multi-species study across the Alps and the Carpathians. *Global Ecol Biog* **18**:78–87.

Thompson KA, Husband BC, Maherali H (2014) Climatic niche differences between diploid and tetraploid cytotypes of *Chamerion angustifolium* (Onagraceae). *Am J Bot* **101**:1868–75.

Tibayrenc M, Kjellberg F, Ayala FJ (1990) A clonal theory of parasitic protozoa: the population structures of *Entamoeba*, *Giardia*, *Leishmania*, *Naegleria*, *Plasmodium*, *Trichomonas*, and *Trypanosoma* and their medical and taxonomical consequences. *Proc Natl Acad Sci USA* **87**:2414–8.

Vandepitte K, Roldán-Ruiz I, Jacquemyn H, *et al.* (2010) Extremely low genotypic diversity and sexual reproduction in isolated populations of the self-incompatible lily-of-the-valley (*Convallaria majalis*) and the role of the local forest environment. *Ann Bot* **105**:769–76.

Vergara M, Cushman SA, Ruiz-Gonzalez A (2017) Ecological differences and limiting factors in different regional contexts: landscape genetics of the stone marten in the Iberian Peninsula. *Landscape Ecol* **32**:1269–83.

Volis S, Ormanbekova D, Shulgina I (2016) Role of selection and gene flow in population differentiation at the edge vs. interior of the species range differing in climatic conditions. *Mol Ecol* **25**:1449–64.

Vranckx G, Jacquemyn H, Mergeay J, *et al.* (2014) The effect of drought stress on heterozygosity-fitness correlations in pedunculate oak (*Quercus robur*). *Ann Bot* **113**:1057–69.

Wookey PA, Robinson CH, Parsons AN, *et al.* (1995) Environmental constraints on the growth, photosynthesis and reproductive development of *Dryas octopetala* at a high Arctic polar semi-desert, Svalbard. *Oecologia* **102**:478–89.

Yardeni G, Tessler N, Imbert E, *et al.* (2016) Reproductive isolation between populations of *Iris atropurpurea* is associated with ecological differentiation. *Ann Bot* **118**:971–82.

Zhao LL, Zhang Y, Wang PC, *et al.* (2016) Morphological and genetic variations of *Sophora davidii* populations originating from different altitudes in the mountains of southwestern China. *Flora* **224**:1–6.

Study 5: Development of SSR markers by 454 sequencing in the endemic species *Gentianella praecox subsp. bohemica* (Gentianaceae)

Applications in Plant Sciences

Applications in Plant Sciences 2017 5(1): 1600114

Primer Note

Development of SSR markers by 454 sequencing in the endemic species *Gentianella praecox* subsp. *bohemica* (Gentianaceae)[1]

Mária Šurinová[2,3,5], Jiří Brabec[4], and Zuzana Münzbergová[2,3]

[2]Institute of Botany of the Czech Academy of Sciences, Zámek 1, 252 43 Průhonice, Czech Republic; [3]Department of Botany, Faculty of Science, Charles University in Prague, Benátská 2, 128 01 Prague, Czech Republic; and [4]Muzeum Cheb, Krále Jiřího z Poděbrad 493/4, 350 11 Cheb, Czech Republic

- *Premise of the study:* Polymorphic microsatellite loci were developed and used to genotype individuals of *Gentianella praecox* subsp. *bohemica* (Gentianaceae), a highly protected taxon in Europe, to study the genetic structure of the remaining populations.
- *Methods and Results:* Thirty-eight primer pairs were successfully amplified; of these, 12 polymorphic microsatellite loci were developed using a 454 sequencing approach and used to genotype 180 individuals of *G. praecox* subsp. *bohemica* from six populations. Allelic richness ranged between one and nine alleles per locus. We detected a high frequency of polyploid individuals (77.8%). The highest average percentage of heterozygous genotypes was identified for samples from the Hroby population (75.5%). All loci can also be amplified in the congeneric species *G. praecox* subsp. *praecox*, *G. amarella* subsp. *amarella*, and *G. obtusifolia* subsp. *sturmiana*.
- *Conclusions:* These markers will provide knowledge on patterns of gene flow and population genetic structure, which is necessary for current protection actions and for effective conservation of this species in the future.

Key words: genotyping; Gentianaceae; *Gentianella praecox* subsp. *bohemica*; microsatellites; polyploidy.

Gentianella praecox (A. Kern. & Jos. Kern.) Dostál ex E. Mayer subsp. *bohemica* (Skalický) Holub (IUCN: e.T161825A5500524) is a strictly biennial herb endemic to the Bohemian Massif, with most populations occurring in the Czech Republic but extending to Bavaria (Germany), Upper and Lower Austria, and Poland. *Gentianella* Moench (Gentianaceae) is a highly diverse and taxonomically complicated genus due to seasonal dimorphism, introgression, and hybridization between closely related species (Winfield et al., 2003; Greimler and Jang, 2007; Plenk et al., 2016). It is expected that *G. praecox* subsp. *bohemica* is tetraploid (Oberdorfer, 1983), but cytotype distribution is unknown. It occurs in seminatural, nutrient-poor grasslands. Strong reduction of population size was recorded during the 20th century, probably due to land-use intensification or abandonment of traditional land use, which led to the disintegration of large habitats and fragmentation of original populations. *Gentianella praecox* subsp. *bohemica* is highly protected in Europe (Annexes II and IV of the Habitats Directive; Council of the European Community, 1992). By using amplified fragment length polymorphism, Königer et al. (2012) studied the genetic structure of 11 *G. praecox* subsp. *bohemica* populations, but this taxon is known from 99 localities (Brabec, 2010). For effective protection of this subspecies, it is necessary to identify populations with high genetic diversity so these populations can be prioritized for protection. Moreover, knowledge about the genetic structure of all remaining populations will reveal patterns of gene flow among populations and the potential for inbreeding depression.

[1]Manuscript received 13 September 2016; revision accepted 27 November 2016.

This work was supported by grants from Iceland, Liechtenstein, and Norway (EHP funds 2009–2014) and by the Ministry of the Environment of the Czech Republic. The authors thank Kristina Plenk and Matthias Kropf for providing *Gentianella praecox* subsp. *praecox* samples and Jana Kadlecová and Eva Ničová for technical assistance.

[5]Author for correspondence: maria.surinova@ibot.cas.cz

doi:10.3732/apps.1600114

METHODS AND RESULTS

***Microsatellite development*—**Total genomic DNA of 14 individuals (two individuals per population collected across the whole distribution range) of *G. praecox* subsp. *bohemica* was extracted from dehydrated leaves using the cetyltrimethylammonium bromide (CTAB) method of Lodhi et al. (1994), with all amounts downscaled 10×. The sequencing facility GenoScreen (Lille, France) was used to prepare libraries and design primers. Extracted DNA was pooled for microsatellite library preparation. The fragmented DNA was hybridized with eight probes (TG, TC, AAC, AAG, AGG, ACG, ACAT, and ACTC) to enrich the DNA library. Sequencing was performed using a GS FLX sequencer (Roche, 454 Life Sciences, Branford, Connecticut, USA). A total of 19,152 reads were obtained. Raw sequencing data were submitted to the National Center for Biotechnology Information (NCBI) Sequence Read Archive (accession no. SRR5113067). QDD software (Meglécz et al., 2009) with default settings was used to identify microsatellite loci and for design of the microsatellite primers. A total of 3017 reads contained microsatellite motifs, and 373 candidate microsatellite loci were identified (Appendix S1), with an average sequence length of 325 bp. Markers belonged to di-, tri-, tetra-, penta-, and hexanucleotide

Applications in Plant Sciences 2017 5(1): 1600114; http://www.bioone.org/loi/apps © 2017 Šurinová et al. Published by the Botanical Society of America. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC-BY-NC-SA 4.0), which permits unrestricted noncommercial use and redistribution provided that the original author and source are credited and the new work is distributed under the same license as the original.

repeats (40.2%, 52.8%, 5.4%, 0.8%, and 0.8%, respectively). Across all candidate loci, 3378 primer pairs (3–15 primer pairs per locus) were designed using Primer3, as implemented within QDD (Malausa et al., 2011) with amplicon lengths ranging between 90 and 319 bp. For each microsatellite candidate locus, one primer pair was selected for further analysis. Of these, we selected 50 primer pairs (Appendix S1) recommended by GenoScreen to identify polymorphic markers. Primers were synthesized (Sigma-Aldrich, St. Louis, Missouri, USA) with M13 tails preceding the 5′ end of the forward primer sequences (Schuelke, 2000). Six individuals from six populations of *G. praecox* subsp. *bohemica* (Appendix 1) were used to test amplification efficiency and polymorphism. DNA amplification was performed in 10-μL reactions consisting of 5 μL of QIAGEN Multiplex PCR Master Mix (QIAGEN, Hilden, Germany), 0.25 μL of each M13-labeled forward, reverse, and fluorolabeled (5′-FAM) M13 primer (10 μM each in initial volume), 20 ng of DNA dissolved in 1 μL TE buffer, and 3.25 μL of H_2O.

The following PCR protocol was performed using an Eppendorf Mastercycler pro S Thermal Cycler (Eppendorf, Hamburg, Germany): an initial denaturation step at 95°C for 15 min; followed by 25 cycles of denaturation (95°C for 20 s), annealing (59°C for 30 s), and extension (72°C for 20 s); followed by 10 cycles of denaturation (95°C for 30 s), annealing (53°C for 45 s), and extension (72°C for 45 s); and a final extension at 72°C for 10 min. Thirty-eight primer pairs (76%) were successfully amplified. Due to allele dosage uncertainty in polyploid individuals, preliminary statistics included determination of polymorphic information content (PIC) for each locus by PICcalc (Nagy et al., 2012). Based on PIC, 20 (53%) of the 38 primer pairs were selected for detailed variability screening on 36 individuals of *G. praecox* subsp. *bohemica* (two individuals from each population). Based on the multiplex PCR performance and variability screening, 12 polymorphic primer pairs were identified.

To confirm primer specificity for these 12 loci, we ran PCRs for each primer pair separately under the same conditions described in the next paragraph. PCR products were purified using the QIAquick PCR Purification Kit (QIAGEN) and cloned using pGEM-T Vector Systems II (Promega Corporation, Madison, Wisconsin, USA) in accordance with the manufacturer's instructions, but downscaled to half reactions. Approximately 10 colonies per sample were transferred into 20 μL of ddH_20 and denatured at 95°C for 10 min. These served as templates for subsequent PCR amplifications for sequencing. Sequencing was performed by the commercial company SEQme (Dobříš, Czech Republic), and the resulting sequences were aligned using MAFFT 7.017 (Katoh et al., 2002) as implemented in Geneious 8.1.6 (Kearse et al., 2012). Repeat motifs with variation in number of repeats were confirmed in the obtained sequences. GenBank accession numbers of identified sequences for 12 loci of *G. praecox* subsp. *bohemica* are provided in Table 1.

***Genotyping*—**Total DNA was extracted from 180 *G. praecox* subsp. *bohemica* individuals from six populations and from 114 individuals from eight populations of three closely related taxa (Appendix 1) for initial primer screening. DNA amplification was carried out in three multiplex reactions consisting of 2.5 μL of QIAGEN Multiplex PCR Master Mix and 10 ng of DNA dissolved in 0.5 μL of TE buffer. For multiplex mix I (MM I), the PCR contained 1.1 μL of primer mix (10 μM each in initial volume) and 0.9 μL of H_2O, for MM II the PCR consisted of 1.1 μL of primer mix (10 μM each in initial volume) and 0.9 μL of H_2O, and for MM III the PCR contained 0.7 μL of primer mix (10 μM each in initial volume) and 1.3 μL of H_2O. The sequence, labeling, motif information, final volumes, and PCR product size range are given in Table 1. The following PCR protocol was performed using an Eppendorf Mastercycler pro S Thermal Cycler: an initial denaturation step at 95°C for 15 min; followed by 35 cycles of denaturation (95°C for 20 s), annealing (59°C for 30 s), and extension (72°C for 20 s); and a final extension at 72°C for 10 min. PCR products were diluted with ddH_2O in these ratios: 1 : 2 (PCR product of MM I and MM II PCRs : ddH_2O), 1 : 9 (PCR product of MM III PCR : ddH_2O). Each PCR product (1 μL) was mixed with 11 μL of a 120 : 1 solution of formamide : size standard (GeneScan 500 LIZ; Thermo Fisher Scientific, Waltham, Massachusetts, USA). Fragment lengths were determined by capillary gel electrophoresis with an ABI 3130 Genetic Analyzer using GeneMapper 4.0 (Thermo Fisher Scientific). Using SPAGeDi (Hardy and Vekemans, 2002), we calculated the number of alleles per locus, which ranged between one and nine (Table 2). All markers were polymorphic in all *G. praecox* subsp. *bohemica* populations, except marker GbM48, which was monomorphic in the Zidkovi population. The highest average percentage of heterozygous genotypes was identified for individuals from the Hroby population (75.5%) and the lowest percentage for individuals from the Zidkovi population (50.5%). We detected a high frequency of polyploid individuals (77.8%). The observed heterozygote excess is likely caused by the fact that the species is tetraploid.

We also tested cross-amplification of these loci in three other *Gentianella* taxa: *G. praecox* subsp. *praecox*, *G. amarella* (L.) Börner subsp. *amarella*, and *G. obtusifolia* (F. W. Schmidt) Holub subsp. *sturmiana* (A. Kern. & Jos. Kern.) Holub. We tested 114 individuals from eight populations (Appendix 1). DNA amplification was carried out in three multiplex reactions as described above. Tests for cross-amplification in the three congeneric taxa resulted in successful amplification of up to seven of the 12 polymorphic loci (Table 2). These results (Table 3) demonstrate that these primer pairs may be of broad utility throughout *Gentianella*.

TABLE 1. Characteristics of 12 polymorphic loci designed for genotyping of *Gentianella praecox* subsp. *bohemica*.

Locus[a]	Primer sequences (5′–3′)	Repeat motif	PCR product size range (bp)	Fluorescent label	Volume of forward primer (μL)	Multiplex	GenBank accession no.
GbM46	F: CAACCACAAGAAGCTTCCAA	$(CTT)_4$	81–129	PET	0.1	I	KX420610
	R: GCATTGCCAACAGATGCAG						
GbM11	F: TGGTTTGATTTCAGACCCTTG	$(TTG)_{16}$	138–180	PET	0.25	I	KX420608
	R: CAGGTTGCCCTACCAAGATG						
GbM34	F: GAAGCGTCCGTTTCAGTTTC	$(TGT)_5$	119–152	NED	0.075	I	KX420611
	R: GCTTAGAGCCCAAGATACCTAGA						
GbM3	F: AGTTGAGAATTGGCCTGGAG	$(GTAT)_5$	134–174	VIC	0.125	I	KX420606
	R: GATGCATTGGAAGCAGGATT						
GbM12	F: ATCAGGCATTGCCATTAAGC	$(AC)_5$	96–108	VIC	0.15	II	KX420604
	R: GAGATTCATAGGTGGCGAGG						
GbM19	F: GGAATTCCTTGTGAAGCCAG	$(GAG)_8$	136–202	VIC	0.225	II	KX420609
	R: TTGCTGCTTCTTTTCCATGA						
GbM38	F: TTTCAAGGTTGCTTTTGGCT	$(AGA)_6$	129–162	NED	0.075	II	KX420612
	R: GCCTTGTGTTAAATTAGTTGCAG						
GbM5	F: CTCCTTCCCTTTTCCCAAAC	$(AG)_8$	158–180	PET	0.1	II	KX420615
	R: GCTTATGTCGCAGTGCAGAA						
GbM2	F: GGGAGAAGCGAGTTCAAAAG	$(GGA)_{13}$	147–180	VIC	0.075	III	KX420607
	R: AAGCTGCTAAACTTCAATACTTGC						
GbM48	F: ACCGAAGGCAGTTTCAACAC	$(GGA)_3$	84–93	NED	0.175	III	KX420613
	R: CCAACAAACTTAGCTACCTTAGCA						
GbM39	F: AACAGAGCAAAAACAAAAACAGG	$(AGA)_8$	79–94	VIC	0.05	III	KX420614
	R: CAAGAAAGCAATGAATCCCC						
GbM43	F: AATCATGTCCAGCTCAGCCT	$(CCT)_4$	158–185	NED	0.05	III	KX420605
	R: GCCGACGTAGAATGTTTGGT						

[a]Optimal annealing temperature was 59°C for all loci.

TABLE 2. Results of initial primer screening of 12 microsatellite loci developed in *Gentianella praecox* subsp. *bohemica* and congeners.

Species	Population code[a]	Primer/Indexes[b]	GbM46	GbM11	GbM34	GbM3	GbM12	GbM19	GbM38	GbM5	GbM2	GbM48	GbM39	GBm43
G. praecox subsp. *bohemica*	HROBY	*A*	3	6	4	7	5	9	4	4	4	4	5	6
		A_{ind}	2.3	2.3	1.43	1.73	2.7	2.8	1.63	2.53	1.43	1.77	1.5	2.6
		% Het	96.7	86.7	40	73.3	100	96.7	60	96.7	43.3	63.3	50	100
	PODVORI	*A*	6	4	4	3	5	5	4	4	5	2	3	5
		A_{ind}	2.3	1.9	1.43	1.47	1.7	1.9	1.27	2.57	1.33	1.2	1.3	2.57
		% Het	86.7	86.7	36.7	46.7	56.7	70	26.7	100	33.3	20	30	100
	POLNA	*A*	4	5	4	4	4	5	4	4	6	3	2	5
		A_{ind}	1.43	2.07	1.83	1.5	2.17	2	1.43	2.47	1.5	1.3	1.13	2.43
		% Het	40	80	73.3	50	83.3	80	43.3	100	50	30	13.3	100
	VYSNY	*A*	7	3	2	3	5	6	3	5	3	3	3	4
		A_{ind}	1.73	1.73	1.5	1.13	2.13	2.63	1.53	2.6	1.5	1.27	1.5	2.23
		% Het	73.3	73.3	50	13.3	83.3	100	53.3	100	50	26.7	50	100
	ZIDKOVI	*A*	5	2	4	6	2	4	2	3	3	1	2	3
		A_{ind}	2.7	1.77	1.1	1.37	1.53	2.3	1.03	2.03	1.13	1	1.17	2.1
		% Het	100	76.7	10	36.7	53.3	100	3.3	100	13.3	0	16.7	96.7
	VANIC	*A*	9	4	3	3	4	5	4	4	3	3	3	4
		A_{ind}	2.8	2.27	1.03	1.53	1.63	2.77	1.67	2.47	1.43	1.53	1.57	2.53
		% Het	96.7	100	6.7	53.3	60	96.7	66.7	100	43.3	46.7	56.7	100
G. amarella subsp. *amarella*	VANIC	*A*	4	3	2	3	4	4	4	3	4	2	3	7
		A_{ind}	1.63	1.67	1.17	1.33	2.17	2.17	1.17	2.17	2	1.83	1.17	2.83
		% Het	60	50	16.7	33.3	100	100	16.7	100	83.3	83.3	16.7	100
	ČER.S.	*A*	4	1	1	1	2	2	1	2	2	2	1	3
		A_{ind}	2.53	1	1	1	2	2	1	2	2	2	1	3
		% Het	96.7	0	0	0	100	100	0	100	100	100	0	100
G. obtusifolia subsp. *sturmiana*	PP PILA	*A*	4	5	4	6	2	7	4	2	2	3	3	4
		A_{ind}	1.43	2.3	1.6	1.5	1.7	2.6	1.6	1.2	1.1	1.4	1.4	2.6
		% Het	43.3	90	60	50	70	90	60	20	10	40	40	100
	KOCEL	*A*	4	4	3	3	4	7	4	3	1	1	2	4
		A_{ind}	1.77	2.5	1.7	1.6	2.3	2.6	1.8	1.8	1	1	1.1	2.8
		% Het	63.3	100	70	60	100	100	80	70	0	0	10	100
	RANK	*A*	5	3	1	2	3	2	1	2	1	1	1	2
		A_{ind}	1.5	1.1	1	1.3	2.3	2	1	2	1	1	1	2
		% Het	50	10	0	30	100	100	0	100	0	0	0	100
G. praecox subsp. *praecox*	BUBE	*A*	6	2	3	4	4	3	2	2	4	5	3	2
		A_{ind}	2.6	1.71	1.21	1.29	1.54	1.71	1.21	2.29	1.25	1.21	1.17	2.46
		% Het	100	62.5	20.8	29.2	45.8	66.7	20.8	100	25	20.8	16.7	100
	GIEE	*A*	1	1	2	2	2	1	1	2	3	3	3	2
		A_{ind}	1.8	1.25	1	1	1.05	1	1.4	2	1	1.25	1.15	2.3
		% Het	75	25	0	0	5	0	40	20	0	25	15	20
	LEOE	*A*	2	2	2	1	2	1	1	2	3	4	2	2
		A_{ind}	1.79	1	1	1.13	1.83	1	1.04	2	1.13	1.04	1.08	2.13
		% Het	79.2	0	0	12.5	83.3	0	4.2	100	12.5	4.2	8.3	100

Note: *A* = number of alleles; A_{ind} = mean number of alleles per individual; % Het = percentage of heterozygous genotypes.

[a] Population code refers to collecting locality. Detailed information is provided in Appendix 1.

[b] Due to allele ambiguity in polyploids, the $H_o : H_e$ ratio is replaced by the percentage of heterozygous genotypes.

TABLE 3. Allele size ranges obtained during cross-amplification trials of microsatellite loci isolated from *Gentianella praecox* subsp. *bohemica* and tested in three additional taxa.

Locus	*G. amarella* subsp. *amarella*	*G. praecox* subsp. *praecox*	*G. obtusifolia* subsp. *sturmiana*
GbM46	111–126	114–117	114–126
GbM11	159–168	150–165	138–168
GbM34	131–134	128–131	119–131
GbM3	142–170	134–158	134–170
GbM12	100–102	100–106	100–108
GbM19	142–154	142–157	142–178
GbM38	137–143	137–146	134–153
GbM5	168–176	168–176	168–178
GbM2	150–177	150–156	147–153
GbM48	84–87	87–90	84–90
GbM39	85–91	82–91	79–85
GbM43	167–179	164–176	167–176

CONCLUSIONS

We developed and successfully multiplexed 12 polymorphic markers in several taxa of *Gentianella*. These polymorphic loci will be valuable for the future management of the extremely rare *G. praecox* subsp. *bohemica*.

LITERATURE CITED

BRABEC, J. 2010. Monitoring hořečku mnohotvarého českého (*Gentianella praecox* subsp. *bohemica*) v ČR (sezóna 2010). [Monitoring of *Gentianella praecox* subsp. *bohemica* in the Czech Republic (vegetation season 2007)]. Deposit in: Central Office of Agentura Ochrany Přírody a Krajiny České Republiky, Prague, Czech Republic.

COUNCIL OF THE EUROPEAN COMMUNITY. 1992. Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora. *Official Journal of the European Communities Series L* 206: 7–49.

GREIMLER, J., AND C. G. JANG. 2007. *Gentianella stiriaca*, a case of reticulate evolution in the northeastern and eastern Central Alps. *Taxon* 56: 857–870.

HARDY, O. J., AND X. VEKEMANS. 2002. SPAGeDi: A versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* 2: 618–620.

KATOH, K., K. MISAWA, K. KUMA, AND T. MIYATA. 2002. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* 30: 3059–3066.

KEARSE, M., R. MOIR, A. WILSON, S. STONES-HAVAS, M. CHEUNG, S. STURROCK, S. BUXTON, ET AL. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics (Oxford, England)* 28: 1647–1649.

KÖNIGER, J., C. A. REBERNIG, J. BRABEC, K. KIEHL, AND J. GREIMLER. 2012. Spatial and temporal determinants of genetic structure in *Gentianella bohemica*. *Ecology and Evolution* 2: 636–648.

LODHI, M. A., G. N. YE, N. F. WEEDEN, AND B. I. REISCH. 1994. A simple and efficient method for DNA extraction from grapevine cultivars and *Vitis* species. *Plant Molecular Biology Reporter* 12: 6–13.

MALAUSA, T., A. GILLES, E. MEGLÉCZ, H. BLANQUART, S. DUTHOY, C. COSTEDOAT, V. DUBUT, ET AL. 2011. High-throughput microsatellite isolation through 454 GS-FLX Titanium pyrosequencing of enriched DNA libraries. *Molecular Ecology Resources* 11: 638–644.

MEGLÉCZ, E., C. COSTEDOAT, V. DUBUT, A. GILLES, T. MALAUSA, N. PECH, AND J. F. MARTIN. 2009. QDD: A user friendly program to choose microsatellite markers and design primers from large sequencing projects. *Bioinformatics (Oxford, England)* 26: 403–404.

NAGY, S., P. POCZAI, I. CERNÁK, A. M. GORJI, G. HEGEDUS, AND J. TALLER. 2012. PICcalc: An online program to calculate polymorphic information content for molecular genetic studies. *Biochemical Genetics* 50: 670–672.

OBERDORFER, E. 1983. Pflanzensoziologische Exkursionsflora. Eugen Ulmer Verlag, Stuttgart, Germany.

PLENK, K., F. GÖD, M. KRIECHBAUM, AND M. KROPF. 2016. Genetic and reproductive characterisation of seasonal flowering morphs of *Gentianella bohemica* revealed strong reproductive isolation and possible single origin. *Plant Biology* 18: 111–123.

SCHUELKE, M. 2000. An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology* 18: 233–234.

WINFIELD, M. O., P. J. WILSON, M. LABRA, AND J. S. PARKER. 2003. A brief evolutionary excursion comes to an end: The genetic relationship of British species of *Gentianella* sect. *Gentianella* (Gentianaceae). *Plant Systematics and Evolution* 237: 137–151.

APPENDIX 1. Accession information for *Gentianella* species used in this study.[a]

Species name	Population code	Collection locality	Country	*n*	Latitude	Longitude
G. praecox (A. Kern. & Jos. Kern.) Dostál ex E. Mayer subsp. *bohemica* (Skalický) Holub	HROBY	Hroby	CZ	30	49.3932222	14.85622
G. praecox subsp. *bohemica*	PODVORI	Podvoří	CZ	30	48.8356111	14.20819
G. praecox subsp. *bohemica*	POLNA	Polná in the Šumava Mountains	CZ	30	48.7928056	14.14997
G. praecox subsp. *bohemica*	VYSNY	Vyšný	CZ	30	48.8266944	14.30211
G. praecox subsp. *bohemica*	ZIDKOVI	Olešnice in the Orlické Mountains	CZ	30	50.3619444	16.28389
G. praecox subsp. *bohemica*	VANIC	Nature Reserve Opolenec	CZ	30	49.0866667	13.79706
G. amarella (L.) Börner subsp. *amarella*	VANIC	Nature Reserve Opolenec	CZ	6	49.0866667	13.79706
G. amarella subsp. *amarella*	ČER.S.	Kouty nad Desnou	CZ	10	50.1225	17.16111
G. obtusifolia (F. W. Schmidt) Holub subsp. *sturmiana* (A. Kern. & Jos. Kern.) Holub	PP PILA	Pila u Karlových Varů	CZ	10	50.1747222	12.92694
G. obtusifolia subsp. *sturmiana*	KOCEL	Kocelovice	CZ	10	49.475	13.82444
G. obtusifolia subsp. *sturmiana*	RANK	Rankovice	CZ	10	50.0067778	12.84208
G. praecox subsp. *praecox*	BUBE	Buchberg, Lower Austria	AU	24	48.376944	15.39722
G. praecox subsp. *praecox*	GIEE	Gießhübl, Lower Austria	AU	20	48.320833	15.36306
G. praecox subsp. *praecox*	LEOE	Leopolds, Lower Austria	AU	24	48.427778	15.28611

Note: AU = Austria; CZ = Czech Republic; *n* = number of individuals.

[a] Because all investigated species are rare and highly protected, it was not possible to sample whole plants for herbarium vouchers. Leaf samples were collected in the field for up to five individuals per population and were dried in silica gel before performing DNA extraction. The leaf samples and DNA extracts were deposited at the Institute of Botany of the Czech Academy of Sciences, Průhonice, Czech Republic.

Study 6: Strong fluctuations in aboveground population size do not limit genetic diversity in populations of an endangered biennial species

Oecologia (2018) 187:863–872
https://doi.org/10.1007/s00442-018-4152-0

CONSERVATION ECOLOGY – ORIGINAL RESEARCH

Strong fluctuations in aboveground population size do not limit genetic diversity in populations of an endangered biennial species

Zuzana Münzbergová[1,2] · Maria Šurinová[1,2] · Iveta Husáková[1,2] · Jiří Brabec[3]

Received: 16 November 2017 / Accepted: 23 April 2018 / Published online: 26 April 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

Assessing genetic diversity within populations of rare species and understanding its determinants are crucial for effective species protection. While a lot is known about the relationships between genetic diversity, fitness, and current population size, very few studies explored the effects of past population size. Knowledge of past population size may, however, improve our ability to predict future population fates. We studied *Gentianella praecox* subsp. *bohemica*, a biennial species with extensive seed bank. We tested the effect of current, past minimal and maximal population size, and harmonic mean of population sizes within the last 15 years on genetic diversity and fitness. Maximum population size over the last 15 years was the best predictor of expected heterozygosity of the populations and was significantly related to current population size and management. Plant fitness was significantly related to current as well as maximum population size and expected heterozygosity. The results suggested that information on past population size may improve our understanding of contemporary genetic diversity across populations. They demonstrated that despite the strong fluctuations in population size, large reductions in population size do not result in immediate loss of genetic diversity and reduction of fitness within the populations. This is likely due to the seed bank of the species serving as reservoir of the genetic diversity of the populations. From a conservation point of view, this suggests that the restoration of small populations of short-lived species with permanent seed bank is possible as these populations may still be genetically diverse.

Keywords Bottleneck effect · Expected heterozygosity · Microsatellites · Seed bank longevity · Storage effect

Introduction

Understanding determinants of genetic diversity is a key prerequisite for its effective conservation. The genetic diversity of a population is determined by a wide range of factors related to species biology, environmental conditions of the locality, as well as the structure of the whole landscape (Frankham et al. 2010). Among others, it has been demonstrated that genetic diversity of a population is strongly linked to its current size and isolation (e.g., Jump and Penuelas 2006; Leimu et al. 2006; Sork and Smouse 2006). While this relationship is usually viewed as the effect of population size on genetic diversity, the relationship is clearly reciprocal as genetic diversity has many significant effects on plant fitness that feed back to modify population size (Ellstrand and Elam 1993; Reed and Frankham 2003).

The importance of current population size for genetic diversity has received a lot of attention in the past. However, recent studies suggested that not only current characteristics of the populations, but also historical characteristics such as past habitat size, past habitat isolation, past habitat continuity, or past management may be important determinants of species diversity, as well as genetic diversity within species (e.g., Lindborg and Eriksson 2004; Metzger et al. 2009; Münzbergová et al. 2013; Husáková and Münzbergová 2014;

Communicated by Monica Geber.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00442-018-4152-0) contains supplementary material, which is available to authorized users.

✉ Zuzana Münzbergová
zuzmun@natur.cuni.cz

[1] Department of Botany, Faculty of Science, Charles University, Prague, Benátská 2, 128 01 Prague 2, Czech Republic

[2] Institute of Botany, Czech Academy of Sciences, Zámek 1, 252 43 Průhonice, Czech Republic

[3] Muzeum Cheb, Krále Jiřího z Poděbrad 493/4, 350 11 Cheb, Czech Republic

Hemrová and Münzbergová 2015; Lehtila et al. 2016; Reisch et al. 2017). This may be caused by time lags in the response of species as well as specific genes to landscape changes (Herben et al. 2006). For populations with a persistent seed bank, this time lag may result in extinction debt (Plue et al. 2017).

Studies directly considering the effect of past population size on current genetic diversity usually deal with the effect of past reduction of the population size, the so-called bottleneck effects (Nei et al. 1975). An increasing number of studies have discussed the importance of bottleneck effects on genetic diversity in animals (e.g., Keller et al. 2001; England et al. 2003; Broquet et al. 2010; Radespiel and Bruford 2014; Tison et al. 2015; Markov et al. 2016; Zenboudji et al. 2016), and less commonly also in plants (e.g., Gonzalez-Astorga et al. 2005; Edwards et al. 2014; Jahnke et al. 2015). Despite the popularity of the topic, most of the studies only indirectly infer the degree of bottleneck from genetic data, without having any direct evidence for its occurrence (but see Leberg 1992 and England et al. 2003 for nice experimental studies on the topic using short-lived animals). The predominant message from these studies is that the restoration of populations, which have experienced strong reductions in size, may be difficult due to a depauperate gene pool. The strength of the bottleneck effect depends on the intensity and duration of the effect with short-term, intense bottlenecks having stronger negative effects, especially on expected heterozygosity of the populations (England et al. 2003). However, low levels of immigration may be effective in mitigating the genetic loss due to the bottleneck (Keller et al. 2001).

Recent studies (Münzbergová et al. 2013; Takkis et al. 2013; Plue et al. 2017; Reisch et al. 2017) demonstrated that past landscape structure may be a better predictor of genetic diversity and plant fitness than current landscape structure. This supports the idea that past events, possibly including bottlenecks, strongly influence patterns of genetic diversity into the future. While these four studies used data on past landscape structure, they did not have data on past population size. By demonstrating that a past, reduced, population size is a better predictor of current genetic diversity than current population size, we could obtain direct evidence that the past events led to the current genetic diversity. While such an approach seems like an obvious direct test of a potential bottleneck effect, we are, in fact, not aware of any study, which attempted to do this. We expect that the absence of such studies is because historical data on population size are only rarely available. Studies which do provide data on past population size have typically been performed with only a few population (e.g., in Van Geert et al. 2008 for *Primula vulgaris* with seven populations and Koniger et al. 2012 for nine populations of *Gentianella praecox* subsp. *bohemica*) or the historical data on population size are available only for short period—2 or 3 years (Fischer and Matthies 1998a; Jacquemyn et al. 2003). In such cases, the data cannot be used for statistical analyses and/or have low predictive power. However, due to the ongoing monitoring programs of many rare species world-wide (in EU, e.g., as a part of Natura 2000 initiative), the number of long-term data sets on population size is likely to rapidly increase in the future.

In this study, we analyzed a unique data set on the population size of 30 populations of a critically endangered biennial species with an extensive seed bank, *Gentianella praecox* subsp. *bohemica*. These populations go through frequent, short-term, bottlenecks. We estimated population size as the number of flowering individuals in each population yearly over the last 15 years. We linked the census data to estimates of genetic diversity and plant fitness from the last year of the study and aimed to answer the following questions: (1) What is the effect of past and present population size on genetic diversity of the populations? (2) What is the effect of population genetic diversity and current and historical population size on plant fitness? As genetic diversity and fitness of plants are also known to be affected by wide range of other factors, such as isolation of the population and quality of the habitat (e.g., Menges and Dolan 1998; Koniger et al. 2012), we additionally asked: (3) What is the effect of isolation of the population and quality of the habitat on the genetic diversity and plant fitness?

We hypothesized that the past population size will be a better predictor of genetic diversity and fitness than the current population size. We also hypothesized that the past population minimum will be a better predictor of genetic diversity and fitness than the past maximum population size. Plant fitness will be primarily related to genetic diversity of the populations and not to population size. Finally, we hypothesized that both genetic diversity and plant fitness will also be related to habitat isolation and management of the localities.

Methods

Studied species

Gentianella praecox subsp. *bohemica* (Skalický) Holub is a biennial plant with an erect, branched stem. The size of the plant and number of flowers in inflorescence vary with season and habitat type. Plants range in height mostly from 5 to 35 cm. The number of flowers per plant is usually 2–50. *Gentianella praecox* subsp. *bohemica* blooms from late August to early October. The seeds ripen during the autumn. A proportion of the seeds germinate the following spring, while the rest are stored in the seed bank (Brabec et al. 2011).

Gentianella praecox subsp. *bohemica* reproduces only by seed (Kirschner and Kirschnerová 2000; Brabec 2010). *Gentianella praecox* subsp. *bohemica* is capable of spontaneous selfing (autogamy), which results in relatively high percentage of developed seeds. Seed production by selfing is on average 5–55% of the seed production from outcrossing (according to several years of experiments—see Reitschläger 2000; Klaudisová 2003). The lack of pollinators can lead to lower viability of the populations. The most common pollinators are likely honey-bees and bumble-bees (Göldel 2012).

Gentianella praecox subsp. *bohemica* is n endemic taxon of the Bohemian Massif and a subendemic species of the Czech Republic. It occurs mainly on current or recently abandoned pastures and on regularly managed meadows. Historically, more than 650 populations were documented in the Czech Republic with only 73 remaining up to know. Most of the current populations had no or very few flowering individuals over the last 5 years (Brabec et al. 2011; Martinec 2015). The loss of the populations in the 20th century is mainly related to changes in the landscape management (especially abandonment of mowing and grazing, eutrophication, shrub encroachment, and reforestation of habitats), which lead to the deterioration of overall conditions for the survival and reproduction of *G. p.* subsp. *bohemica* (Brabec 2010).

Population data collection

For this study, we used 30 populations of *G. p.* subsp. *bohemica* coming from all five major areas of occurrence in the Czech Republic (see Fig. 1). As individual populations differed in their sizes (number of flowering individuals) greatly, we decided to collect the leaf samples for genetic analysis in all populations with at least ten flowering individuals (i.e., 30 populations in the Czech Republic), but sometimes not all plants could be used for sampling (due to low quality of the leaves).

For all the studied populations, population size was expressed as the number of flowering individuals per population in the present and in the past. For 27 populations, we had data on population size over the whole 15-year period, between 2000 and 2014. For the remaining three, we had data on population size from fewer years (4–12 years). This is because these populations were discovered just after this study was initiated. In our analyses (as described below), we compared results with and without these three populations. As these results were largely similar, we only present results based on the complete data set.

For all populations, we collected data on metrics of plant performance in the last year of the study. Specifically, we have data on total plant height, total number of capsules, and number of developed and undeveloped seeds in the upper most capsule. These data were collected for approximately

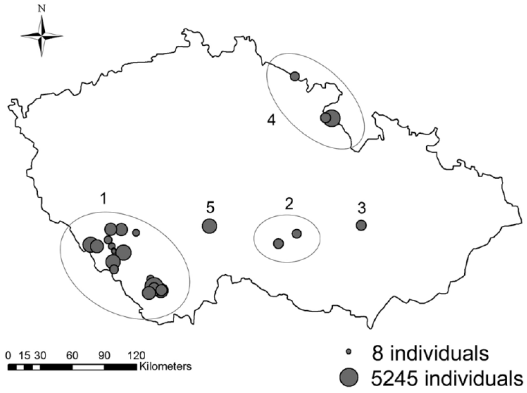


Fig. 1 All studied localities of *Gentianella praecox* subsp. *bohemica* in the Czech Republic. The localities are distributed across five regions of its distribution—(1) the Šumava mountains and the Šumavsko-novohradské foothills; (2) the Bohemian-Moravian highlands; (3) the Drahanská highland; (4) north-eastern Bohemia in the Orlické foothills, on the Broumovsko region and in the Giant mountains; (5) the Táborsko region with only one isolated locality

the same number of individuals as the genetic data (see below), and most often, the same individuals were sampled in each data set. Sampling the same individual for performance and genetic data sets was, however, not always possible, because fruiting plants cannot always be sampled for genetics (some already do not have any leaves) and the flowering plants suitable for genetic sampling cannot always be collected for seed production (their seeds are not yet ripe). We thus only used population averages of the performance data in all the subsequent tests.

Habitat conditions

To describe habitat conditions at the sites, we used one vegetation relevé (5 × 5 m) collected for a purpose of another study at each locality (Křenová et al. in prep.). The same study also provided data on quality of the management since 1999. It was scored using a scale from 1 to 5 with 1 indicating optimal management and 5 indicating management which is missing or has negative impact on *G. p.* subsp. *bohemica* populations, such as mowing at the time of flowering. For this study, we used information on the management in the period 2011–2016, i.e., in the period including the year of genetic sample collection. We also extracted the highest management score (indicating the least suitable management) for each population since 1999. As the least suitable management did not have any significant effect on any of the response variables ($p \geq 0.48$ in all cases), only the current management is presented further for simplicity.

To estimate isolation of each population, we considered all populations within the radius of 50 km from each target population that existed any time between 2002 and 2014. We used this information to calculate habitat isolation using modification of the approach of Knappová et al. (2012). Specifically, we calculated the isolation of a population as follows:

$$I_j = -\log \sum_{k=1}^{n} \left(A_k / D_{jk}^2 \right), j \neq k,$$

where I_j is the isolation of a population j; k represents all the surrounding populations within a 50-km distance of population j. D_{jk} is the distance between populations j and k. A_k was defined in two different ways. First, it was set to 1 in all cases, indicating just population presence during the period of time. In the second version, it was set to population size in 2014 to account for current size differences between populations. As the former version of the isolation index did not have any significant effect on any of the response variables ($p \geq 0.45$ in all cases), only the latter is presented further for simplicity.

Plant material

We collected the samples for the genetic analyses from 30 individuals per population in autumn 2014. When there were fewer than 30 flowering plants with viable leaves, we sampled the maximum number of plants possible. This happened in 11 populations with 6–21 individuals sampled per population (mean 13, Online Resource 1). These plants largely match the plants studied for the fitness (with the exceptions stated above), so the two data sets are comparable. Leaf samples were dried in silica gel. The leaf samples and DNA extracts are deposited at the Institute of Botany of the Academy of Sciences of the Czech Republic.

Microsatellite analyses

Total genomic DNA was extracted from dehydrated leaves using a CTAB method (Lodhi et al. 1994), but 10× downscaled. Quality and quantity of the extracted DNA were measured by NanoDrop 2000 (Thermo Scientific, Waltham, USA), and all the samples were normalized to 20 ng/µl for subsequent PCR. We used 12 polymorphic microsatellite markers previously developed by Šurinová et al. (2017). We used the protocol described in Šurinová et al. (2017) to perform the analyses. The primary data are shown in Online Resource 2.

Data analyses

We used the genetic data to express observed (H_o) and expected heterozygosity (H_e) of the populations. In addition, we calculated rarefied allelic richness to account for the fact that allelic richness is a function of population size. All these calculations were done using SPAGeDi 1.5. (Hardy and Vekemans 2002). As H_o was ≥ 0.99 in all populations, these data were not analyzed and presented further. We eventually also excluded the rarified allelic richness as it was highly correlated to expected heterozygosity ($r = 0.99$), and expected heterozygosity is more commonly used in studies on plant population genetics and is also expected to better reflect the bottleneck effects (England et al. 2003).

The population size strongly varied over time. Thus, we used population size in the year of material collection (2014, further referred to as current population size), minimum and maximum of population size between 2000 and 2014 and harmonic mean population size over the observation period as predictors of genetic diversity and fitness (see Fig. 2a for comparison of these values). We used population minimum as an indicator of the intensity of the bottleneck in the populations. In contrast, we expected that the population maximum may reflect the potential of the population to profit from the suitable conditions. Harmonic mean of the population size was used, because it was suggested to well reflect effective population size in species with highly fluctuating populations in the previous studies (Hartl and Clark 2007) including study using *G. p.* subsp. *bohemica* (Koniger et al. 2012). All the measures of population size were log transformed (minimum was log + 1 transformed) for the analyses.

We used detrended correspondence analysis (DCA), a unimodal indirect multivariate analysis (Lepš and Šmilauer 2003), to analyze the main gradients in the vegetation of the localities. We used position of each vegetation relevé on the first and the second ordination axes to express summarized vegetation composition of the locality. The multivariate analysis was done using CANOCO 5 (ter Braak and Šmilauer 1998). As the vegetation data did not have any significant effect on any of our response variables ($p \geq 0.22$ in all cases), they are not presented further for simplicity.

To assess the determinants of the genetic diversity of the populations, we used simple linear regression and tested the effect of logarithm of minimum, maximum, current and harmonic mean population size, population isolation, and management on expected heterozygosity. In addition, we tested the effect of all the above population and genetic characteristics on the number of capsules per plant and the number of developed seeds per the upper most capsule. We also considered to use stepwise selection analyses to select the variables best explaining the dependent variables. However, as our predictors were partly intercorrelated and the results thus partly depend on the exact procedure (forward,

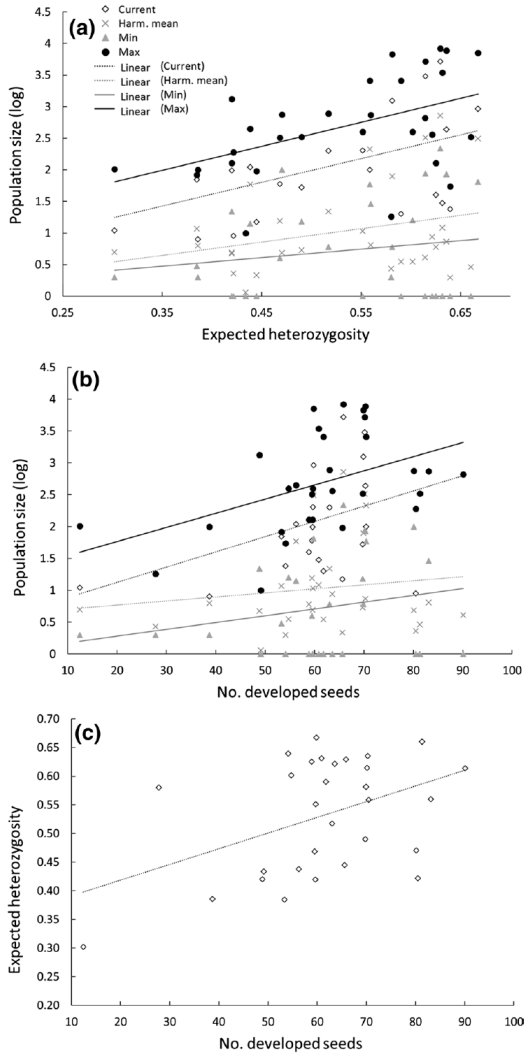


Fig. 2 Effects of population size [Current, minimum (min), maximum (max), and harmonic mean (harm. mean)] on (**a**) expected heterozygosity of the populations (on *x*-axis) and **b** number of developed seeds per capsule (on *x*-axis). **c** Effects of expected heterozygosity of the populations (on *y*-axis) on number of developed seeds (per capsule). Each data point represents a single population. Black symbols and lines indicate significant relationships ($p \leq 0.05$), while grey symbols and lines show relationships which are not significant ($p > 0.05$). Linear (current) indicates the line fitted to the data on current population size, and similarly for the other variables on the *y*-axis

backward) and also on the order of introducing the predictors, we decided to only present the analyses for each predictor separately. To simplify presentations of all the results, all the patterns are shown in a form of a pair-wise correlation matrix. The provided *r* values correspond to square root of the r^2 values obtained from the regression.

Due to fewer than 30 samples in 11 out of 30 of our populations, we repeated the above analyses after accounting for sample size of each population. As none of our dependent variables is related to sample size by its definition, we present these results just for completeness in the Online Resource 3. All the analyses were done in R 3.2.1 (R Development Core Team 2011).

Results

Expected heterozygosity was significantly positively correlated to the current population size and the maximum population size, with the maximum population size showing stronger relationship (Fig. 2a, Table 1). Expected heterozygosity was also significantly positively related to the quality of the management at the localities and marginally significantly decreased with increasing isolation (Table 1).

The number of developed seeds per capsule was significantly related to expected heterozygosity, and current and maximum population sizes (Fig. 2b, c, Table 1). Current population size showed the highest correlation in this case, thought the differences were low (Table 1). In contrast, a number of capsules per plant were positively marginally significantly related only to the minimum population size (Table 1).

Discussion

While the previous studies have explored the relationship between the genetic diversity and the current population size (reviewed in Leimu et al. 2006), we know very little about the relationship between the genetic diversity and the past population sizes in nature. In line with our hypothesis, we demonstrated that the past population size can be a useful predictor of the genetic diversity of the populations. In contrast to our hypothesis, we, however, demonstrated that the maximum population size over the last 15 years is a better predictor of the population genetic diversity than the population minimum and the population harmonic mean. Genetic diversity was also significantly related to the current population size (i.e., in the year of genetic data collection) as expected based on the previous studies. In line with the prediction, fitness measured as seed production was correlated with genetic diversity, but the current and the maximum population size were better predictors of plant fitness than the genetic diversity. Genetic diversity was also significantly related to the current management of the localities and marginally significantly to the habitat isolation. Not surprisingly, all the measures of population size were significantly

Table 1 Pairwise correlation matrix showing correlation coefficient (r) between all variables used in the study

	Current	Harm. mean	Min	Max	Isolation	Manag.	He	Dev. seeds	No. capsules
Current	–	**0.59**	**0.68**	**0.7**	**−0.57**	**−0.54**	**0.46**	**0.47**	0.23
Harm. mean	**0.59**	–	**0.66**	**0.7**	−0.21	−0.08	0.29	0.14	0.04
Min	**0.68**	**0.66**	–	**0.64**	−0.16	−0.25	0.18	0.22	*0.35*
Max	**0.7**	**0.7**	**0.64**	–	−0.26	−0.26	**0.49**	**0.46**	0.21
Isolation	**−0.57**	−0.21	−0.16	−0.26	–	0.4	*−0.33*	−0.25	−0.05
Manag.	**−0.54**	−0.08	−0.25	−0.26	**0.4**	–	**−0.4**	−0.23	−0.05
He	**0.46**	0.29	0.18	**0.49**	−0.33	**−0.4**	–	**0.43**	0.17
Dev. seeds.	**0.47**	0.14	0.22	**0.46**	−0.25	−0.23	**0.43**	–	**0.4**
No. capsules	0.23	0.04	*0.35*	0.21	−0.05	−0.05	0.17	**0.4**	–

It shows the relationship between population size, genetic diversity, and plant fitness. Significant values ($p \leq 0.05$) are shown in bold. Values in italics are marginally significant ($0.05 < p \leq 0.1$). Current, harm. mean, min, and max indicate current, harmonic mean, minimum, and maximum population size between 2000 and 2014. Isolation indicates habitat isolation. Manag. indicates current management at the localities. He indicates expected heterozygosity. Dev. seeds indicates number of developed seeds in the upper most capsule. No. capsules indicates number of capsules per plant

correlated to each other. Despite this, genetic diversity was significantly related only to the current and the maximum population size, but not to the population minimum and/or to its harmonic mean. The absence of this relationship is discussed below.

A significant effect of the past population size, expressed as the number of flowering individuals, on the genetic diversity of the populations may suggest the existence of a time lag in the genetic composition of the populations. The expected time lag in the system is in line with the previous studies demonstrating that the genetic diversity may often experience time lag and the past landscape structure may thus be a good predictor of the genetic diversity of the populations (e.g., Prentice et al. 2006; Münzbergová et al. 2013; Rosengren et al. 2013; Epps and Keyghobadi 2015; Plue et al. 2017; Reisch et al. 2017). The previous studies linking the genetic diversity and the population history, however, dealt only with the past landscape structure and not with the past population size as such data were not available. Clearly, however, also the current population size was an important determinant of the genetic diversity in our system. This result should be interpreted with some caution though, because population size in *G. p.* subsp. *bohemica* is highly variable between years and the significant correlation we observed may be in part dependent on the specific year in which sampling occurred. As a result, we expect that effects of current population size are more dependent on the specific year of the study, while the historical maximum may be a more robust predictor of the genetic diversity independent of the current conditions. In the year of study (2014), the correlation of the current and the maximum population size was in fact quite high ($r = 0.7$), though the correlations tend to be much lower in the other years (not shown).

Based on the wealth of studies detailing the effects of bottlenecks on genetic diversity (e.g., Edwards et al. 2014; Jahnke et al. 2015; Zenboudji et al. 2016), it could be expected that the population minimum will be the best predictor of the genetic diversity of our populations. Strong bottleneck effects are likely to be detected when reductions in population size are especially large and mainly affect expected heterozygosity (England et al. 2003). Populations of *G. praecox* frequently experience large reductions in the number of flowering individuals, exhibiting bottleneck-like behavior. The absence of a bottleneck effects in our system supports the previous studies, indicating that the population bottlenecks are not a universal phenomenon (Ilves et al. 2015).

The most likely explanation for the lack of the bottleneck effect and of the maximum population size being one of the best predictors of the genetic diversity in the system are the very high fluctuations in the population size together with the existence of the extensive seed bank (Brabec et al. 2011). Thanks to this, the maximum population size over the last 15 years may be the best predictor of the current effective population size. We expect that, during population maxima, high quantity of genetically very diverse seeds is stored in the seed bank. These seeds then germinate over an extended period and even during the times when the populations get very small, the seed bank ensures that these populations can be, in fact, very genetically diverse.

To confirm this, it would be useful to study genetic diversity directly in the seed bank of the populations of different current size as well as of different history. While we in fact attempted to do this by sampling and cultivating 1800 seed bank samples, this task turned out to be very challenging due to the strong dormancy and small size of the seeds. Due to this, the number of seedlings obtained was too low for any reasonable analysis. We thus hope to be able to obtain these results at some time in the future. The previous studies exploring the genetic diversity in the seed banks of

other systems (e.g., Mandák et al. 2006, 2012; Mandák and Plačková 2009), however, confirmed that genetic diversity in the seed bank can be indeed very high and may exceed that in the aboveground. Plue et al. (2017) in fact directly demonstrated that seed bank may at least partly contribute to genetic extinction debt of the populations.

The high predictive power of the past maximum population size for genetic diversity may also be related to the mixed-mating system with a variable, but not studied, proportion of autonomous selfing. The selfing rate might have been the lowest in the years with the highest aboveground population size, resulting in a net addition of the genetic diversity to the seed bank, whereas this addition is much lower in the years with a smaller population size and a higher selfing rate. In line with this, Moeller and Geber (2005) and Kennedy and Elle (2008), indeed, suggested that population size may be an important determinant of the selfing rate of a population. The positive effect of the maximum population size on the genetic diversity may be further enhanced by the higher seed production and thus higher addition to the seed bank in the years of the population maxima. Such a pattern, referred to as the Allee effect, has, indeed, been demonstrated in a range of the previous studies (e.g., Groom 1998; Ruane et al. 2014; Anic et al. 2015).

Lack of the bottleneck effect in *G. p.* subsp. *bohemica* suggests that the strong reduction in the population size does not represent a serious problem for *G. p.* subsp. *bohemica* and restoration of the populations is sensible even in populations that are currently very small. This finding thus supports the conclusion of a previous population study on the same system (Bucharová et al. 2012), suggesting that the populations of *G. p.* subsp. *bohemica* can be reasonably restored even if no flowering individuals have been detected in the populations over a period of up to 4 years. The absence of significant effect of the minimum population size likely explains also the lack of effect of the harmonic mean as it incorporates the information on the minimum. Still, this pattern was not expected as the harmonic mean proved to be a significant predictor of the genetic diversity of the populations of *G. p.* subsp. *bohemica* in a study using only few populations (Koniger et al. 2012). Their study, however, involved only larger populations and the fluctuations in size were not as large as in our data set.

Genetic diversity of the populations was also affected by the current population size, a pattern known from a range of the previous studies (reviewed in Leimu et al. 2006, later e.g., Dostálek et al. 2010; Šmídová et al. 2011; Bucharová and Münzbergová 2012; Ilves et al. 2015; Putz et al. 2015; Busch and Reisch 2016). It may be explained by the close feedback between the population size and the genetic diversity (Ellstrand and Elam 1993). Such a relationship may be caused also by differences in the habitat quality affecting plant fitness, population size, and subsequently genetic diversity of the populations (Leimu et al. 2006). This explanation was supported by the fact that genetic diversity was significantly related to the quality of the current management at the localities. Quality of the current management was also significantly related to the current population size, but not to any other population size measures. The observed effects of the maximum population size are thus unaffected by its correlation with the current management. Genetic diversity was also marginally significantly related to the isolation of the populations (e.g., Ellstrand and Elam 1993; Gonzalez-Astorga et al. 2004; Honnay et al. 2007; Heinicke et al. 2016), though the pattern is not ubiquitous (e.g., Young et al. 1999).

In line with numerous previous studies (e.g., Ellstrand and Elam 1993; Leimu et al. 2006; Münzbergová and Plačková 2010; Ensslin et al. 2011), plant fitness measured as the number of developed seeds was significantly related to the genetic diversity. It was also significantly related to the current and the maximum population size, the two significant predictors of the genetic diversity. In contrast to this, the number of capsules per plant was marginally significantly related only to the minimum population size and not to any other characteristics of the populations. The discrepancy between the predictors of the number of seeds per capsule and the number of capsules per plant may be caused by different processes driving these two variables. The number of seeds per capsule is likely the function of pollen quality and quantity (e.g., Wolf and Harrison 2001). In contrast, the number of capsules most likely reflects the genetic constitution of the individual, including the level of inbreeding that may be increased in the populations undergoing strong bottlenecks (Ellstrand and Elam 1993). This possibility was not supported by our data as the expected heterozygosity was independent of the minimum population size in our system. It is, however, possible that the populations, which underwent strong reduction in size, would be detected asinbred in case we used a larger set of genetic markers or genetic markers linked to functional genes.

The fitness measures used in our study were variables related to seed production. While these measures are commonly used to describe plant fitness (see e.g., Leimu et al. 2006), it is obvious that these fitness measures are not perfect. To properly describe plant fitness in natural conditions, we should also consider the ability of the developed seeds to germinate, establish, and produce reproducing individuals as seed production is indeed not a good measure of growth of natural populations (e.g., Heinken-Šmídová and Münzbergová 2012; Dostálek and Münzbergová 2013; van der Meer et al. 2014). However, because seed germination and seedling establishment at natural localities mainly reflect abiotic conditions of the habitats, including this information into the fitness measures would weaken our ability to link the plant fitness to the genetic diversity and the population

size. Thus, while our fitness data cannot be directly used to make inferences about the future performance of the populations, they are a useful proxy of the population ability to grow given that the habitat conditions are suitable.

In line with a range of other studied studies on plants (e.g., Fischer and Matthies 1998a; Leimu and Mutikainen 2005; Leimu 2010; Reisch et al. 2018), we used the number of flowering individuals per population as an estimate of the population size. Such an estimate is used in many programs of monitoring of rare species including the Natura 2000 monitoring program (e.g., Ryttäri et al. 2003; Brabec 2010; Turoňová 2015) and it is easy to do. Fischer and Matthies (1998b) used also an alternative measure of the population size, the area colonized by the species, for *Gentianella germanica,* and they found that these two measures were highly correlated. However, clearly, the number of flowering plants is not the best possible measure of population size as plant populations are composed of seedlings, older non-reproductive plants, and also the seed bank. Moreover, it is known that flowering individuals may constitute only a small fraction of population size in populations of many species (e.g., Dostálek and Münzbergová 2013; van der Meer et al. 2014) and the proportion of flowering plants may strongly vary across populations of the same species (e.g., Oostermeijer et al. 1996; Heinken-Šmídová and Münzbergová 2012; van der Meer et al. 2014). As a result, the conclusions based on the relationship between the population size measured as the number of flowering individuals, the genetic diversity, and the fitness need to be interpreted with caution. Nevertheless, such a measure is often the only available measure of the population size and is thus used in this study and likely will be used also in the future studies on the topic.

A clear weak point of our study is that most of our predictors are significantly related to each other and it is thus impossible to separate their effects. This was also the reason why we decided not to use any stepwise model selection procedure and rather present only the pair-wise relationships. While this may seem as a weak approach to data analyses, we decided to use it as the results are easy to present and are not affected by any arbitrary decision on how exactly the data should be analyzed.

Overall, the study demonstrated that the data on the past population size expressed as the number of flowering individuals can provide important insights into the factors structuring genetic diversity within populations and thus provide useful guidelines for effective protection of the species. In our system, the results suggest that despite the strong fluctuations in the population size of *G. p.* subsp. *bohemica*, the strong declines in population size do not mean loss of genetic diversity and reduction of fitness within the populations expressed as number of seeds per capsule. This is likely due to the extensive seed bank in the populations. From conservation point of view, the results thus suggest that restoration of small populations of short-lived species with extensive seed bank is possible as strong reduction in population size does not mean immediate loss of genetic diversity and fitness of the populations.

Acknowledgements We thank Population Ecology Department (POPEKOL) discussion group and two anonymous reviewers for helpful comments on the manuscript. Data on population size come from monitoring financed by Agency of nature and landscape protection of the Czech Republic (AOPK) to JB. This study was supported by European Economic Area (EEA) funds 2009–2014 and the Ministry of the Environment of the Czech Republic (www.eeagrants.cz, http://www.mzp.cz/).

Author contribution statement ZM and JB conceived and designed the experiments, JB and IH collected the plant material, MŠ performed the experiments, MŠ and ZM analyzed the data, IH created the maps, ZM wrote the manuscript; other authors provided editorial advice.

Compliance with ethical standards

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

References

Anic V, Henriquez CA, Abades SR, Bustamante RO (2015) Number of conspecifics and reproduction in the invasive plant *Eschscholzia californica* (Papaveraceae): is there a pollinator-mediated Allee effect? Plant Biol 17:720–727. https://doi.org/10.1111/plb.12293

Brabec J (2010) http://www.zachranneprogramy.cz/index.php?docId=6145&parentId=6140&spec=rostliny

Brabec J, Bucharová A, Štefánek M (2011) Vliv obhospodařování na životní cyklus hořečku mnohotvarého českého (*Gentianella praecox* subsp. *bohemica*). Příroda 31:85–109

Broquet T et al (2010) Genetic bottlenecks driven by population disconnection. Conserv Biol 24:1596–1605. https://doi.org/10.1111/j.1523-1739.2010.01556.x

Bucharová A, Münzbergová Z (2012) Gene flow among populations of two rare co-occurring fern species differing in ploidy level. PLoS ONE. https://doi.org/10.1371/journal.pone.0045855

Bucharová A, Brabec J, Münzbergová Z (2012) Effect of land use and climate change on the future fate of populations of an endemic species in central Europe. Biol Conserv 145:39–47. https://doi.org/10.1016/j.biocon.2011.09.016

Busch V, Reisch C (2016) Population size and land use affect the genetic variation and performance of the endangered plant species *Dianthus seguieri* ssp *glaber*. Conserv Genet 17:425–436. https://doi.org/10.1007/s10592-015-0794-1

Dostálek T, Münzbergová Z (2013) Comparative population biology of critically endangered *Dracocephalum austriacum* (Lamiaceae) in two distant regions. Folia Geobot 48:75–93. https://doi.org/10.1007/s12224-012-9132-2

Dostálek T, Münzbergová Z, Plačková I (2010) Genetic diversity and its effect on fitness in an endangered plant species, *Dracocephalum austriacum* L. Conserv Genet 11:773–783. https://doi.org/10.1007/s10592-009-9879-z

Edwards CE, Lindsay DL, Bailey P, Lance RF (2014) Patterns of genetic diversity in the rare Erigeron lemmoni and comparison with its more widespread congener, *Erigeron arisolius* (Asteraceae). Conserv Genet 15:419–428. https://doi.org/10.1007/s10592-013-0549-9

Ellstrand NC, Elam DR (1993) Population genetic consequences of small population-size—implications for plant conservation. Annu Rev Ecol Syst 24:217–242. https://doi.org/10.1146/annurev.es.24.110193.001245

England PR, Osler GHR, Woodworth LM, Montgomery ME, Briscoe DA, Frankham R (2003) Effects of intense versus diffuse population bottlenecks on microsatellite genetic diversity and evolutionary potential. Conserv Genet 4:595–604. https://doi.org/10.1023/a:1025639811865

Ensslin A, Sandner TM, Matthies D (2011) Consequences of ex situ cultivation of plants: genetic diversity, fitness and adaptation of the monocarpic *Cynoglossum officinale* L. in botanic gardens. Biol Conserv 144:272–278. https://doi.org/10.1016/j.biocon.2010.09.001

Epps CW, Keyghobadi N (2015) Landscape genetics in a changing world: disentangling historical and contemporary influences and inferring change. Mol Ecol 24:6021–6040. https://doi.org/10.1111/mec.13454

Fischer M, Matthies D (1998a) Effects of population size on performance in the rare plant *Gentianella germanica*. J Ecol 86:195–204. https://doi.org/10.1046/j.1365-2745.1998.00246.x

Fischer M, Matthies D (1998b) RAPD variation in relation to population size and plant fitness in the rare *Gentianella germanica* (Gentianaceae). Am J Bot 85:811–819. https://doi.org/10.2307/2446416

Frankham R, Ballou J, Briscoe D (2010) Introduction to conservation genetics. Cambridge University Press, Cambridge

Göldel B (2012) Reproduction biology: floral visitors, floral scent and reproductive success of the Bohemian gentian *Gentianella bohemica*. Master thesis, Universität Bayreuth, Germany

Gonzalez-Astorga JG, Cruz-Angon A, Flores-Palacios A, Vovides AP (2004) Diversity and genetic structure of the Mexican endemic epiphyte *Tillandsia achyrostachys* E. Morr. ex Baker var. *achyrostachys* (Bromeliaceae). Ann Bot 94:545–551. https://doi.org/10.1093/aob/mch171

Gonzalez-Astorga J, Vovides AP, Cruz-Angon A, Octavio-Aguilar P, Iglesias C (2005) Allozyme variation in the three extant populations of the narrowly endemic cycad *Dioon angustifolium* Miq. (Zamiaceae) from North-eastern Mexico. Ann Bot 95:999–1007. https://doi.org/10.1093/aob/mci106

Groom MJ (1998) Allee effects limit population viability of an annual plant. Am Nat 151:487–496. https://doi.org/10.1086/286135

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. https://doi.org/10.1046/j.1471-8286.2002.00305.x

Hartl D, Clark A (2007) Principles of population genetics. Sinauer Associates Inc., Sunderland

Heinicke S et al (2016) Fragmentation and environmental constraints influence genetic diversity and germination of *Stipa pennata* in natural steppes. Flora 224:42–49. https://doi.org/10.1016/j.flora.2016.06.003

Heinken-Šmídová A, Münzbergová Z (2012) Population dynamics of the endangered, long-lived perennial species, *Ligularia sibirica*. Folia Geobot 47:193–214. https://doi.org/10.1007/s12224-011-9116-7

Hemrová L, Münzbergová Z (2015) The effects of plant traits on species' responses to present and historical patch configurations and patch age. Oikos 124:437–445. https://doi.org/10.1111/oik.01130

Herben T, Münzbergová Z, Mildén M, Ehrlén J, Cousins SAO, Eriksson O (2006) Long-term spatial dynamics of *Succisa pratensis* in a changing rural landscape: linking dynamical modelling with historical maps. J Ecol 94:131–143. https://doi.org/10.1111/j.1365-2745.2005.01063.x

Honnay O, Adriaens D, Coart E, Jacquemyn H, Roldan-Ruiz I (2007) Genetic diversity within and between remnant populations of the endangered calcareous grassland plant *Globularia bisnagarica* L. Conserv Genet 8:293–303. https://doi.org/10.1007/s10592-006-9169-y

Husáková I, Münzbergová Z (2014) Relative importance of current and past landscape structure and local habitat conditions for plant species richness in dry grassland-like forest openings. PLoS ONE. https://doi.org/10.1371/journal.pone.0097110

Ilves A, Metsare M, Tali K, Kull T (2015) The impact of recent colonization on the genetic diversity and fine-scale genetic structure in *Orchis militaris* (L.). Plant Syst Evol 301:1875–1886. https://doi.org/10.1007/s00606-015-1200-7

Jacquemyn H et al (2003) Effects of agricultural land use and fragmentation on genetics, demography and population persistence of the rare *Primula vulgaris*, and implications for conservation. Belg J Bot 136:5–22

Jahnke M, Serra IA, Bernard G, Procaccini G (2015) The importance of genetic make-up in seagrass restoration: a case study of the seagrass *Zostera noltei*. Mar Ecol Prog Ser 532:111–122. https://doi.org/10.3354/meps11355

Jump AS, Penuelas J (2006) Genetic effects of chronic habitat fragmentation in a wind-pollinated tree. Proc Natl Acad Sci USA 103:8096–8100. https://doi.org/10.1073/pnas.0510127103

Keller LF et al (2001) Immigration and the ephemerality of a natural population bottleneck: evidence from molecular markers. Proc R Soc B 268:1387–1394. https://doi.org/10.1098/rspb.2001.1607

Kennedy BF, Elle E (2008) The reproductive assurance benefit of selfing: importance of flower size and population size. Oecologia 155:469–477. https://doi.org/10.1007/s00442-007-0924-7

Kirschner J, Kirschnerová L (2000) *Gentianella* MOENCH—hořeček. In: Slavík B (ed) Květena ČR, vol 6. Academia, Praha, pp 82–98

Klaudisová M (2003) Studium vybraných fází životního cyklu *Gentianella praecox* subsp. *bohemica* a *G. amarella* subsp. *amarella*. Master thesis, Department of Botany, Faculty of Science, Charles University, Prague, Czech Republic

Knappová J, Hemrová L, Münzbergová Z (2012) Colonization of central European abandoned fields by dry grassland species depends on the species richness of the source habitats: a new approach for measuring habitat isolation. Landsc Ecol 27:97–108. https://doi.org/10.1007/s10980-011-9680-5

Koniger J, Rebernig CA, Brabec J, Kiehl K, Greimler J (2012) Spatial and temporal determinants of genetic structure in *Gentianella bohemica*. Ecol Evol 2:636–648. https://doi.org/10.1002/ece3.211

Leberg PL (1992) Effects of population bottlenecks on genetic diversity as measured by allozyme electrophoresis. Evolution 46:477–494. https://doi.org/10.2307/2409866

Lehtila K, Dahlgren JP, Garcia MB, Leimu R, Syrjanen K, Ehrlén J (2016) Forest succession and population viability of grassland plants: long repayment of extinction debt in *Primula veris*. Oecologia 181:125–135. https://doi.org/10.1007/s00442-016-3569-6

Leimu R (2010) Habitat quality and population size as determinants of performance of two endangered hemiparasites. Ann Bot Fenn 47:1–13

Leimu R, Mutikainen P (2005) Population history, mating system, and fitness variation in a perennial herb with a fragmented distribution. Conserv Biol 19:349–356. https://doi.org/10.1111/j.1523-1739.2005.00480.x

Leimu R, Mutikainen P, Koricheva J, Fischer M (2006) How general are positive relationships between plant population size, fitness and genetic variation? J Ecol 94:942–952. https://doi.org/10.1111/j.1365-2745.2006.01150.x

Lepš J, Šmilauer P (2003) Multivariate analysis ecological data using CANOCO. Cambridge University Press, Cambridge

Lindborg R, Eriksson O (2004) Historical landscape connectivity affects present plant species diversity. Ecology 85:1840–1845. https://doi.org/10.1890/04-0367

Lodhi MA, Ye GN, Weeden NF, Reisch BI (1994) A simple and efficient method for DNA extraction from grapevine cultivars and *Vitis* species. Plant Mol Biol Rep 12:6–13

Mandák B, Plačková I (2009) How does population genetic diversity change over time? An experimental seed bank study of *Atriplex tatarica* (Chenopodiaceae). Flora 204:423–433. https://doi.org/10.1016/j.flora.2008.05.005

Mandák B, Bímová K, Mahelka V, Plačková I (2006) How much genetic variation is stored in the seed bank? A study of *Atriplex tatarica* (Chenopodiaceae). Mol Ecol 15:2653–2663. https://doi.org/10.1111/j.1365-294X.2006.02953.x

Mandák B, Zákravský P, Mahelka V, Plačková I (2012) Can soil seed banks serve as genetic memory? A study of three species with contrasting life history strategies. PLoS ONE. https://doi.org/10.1371/journal.pone.0049471

Markov G, Zhelev P, Ben Slimen H, Suchentrunk F (2016) Population genetic data pertinent to the conservation of *Bulgarian chamois* (*Rupicapra rupicapra balcanica*). Conserv Genet 17:155–164. https://doi.org/10.1007/s10592-015-0768-3

Martinec P (2015) Hořeček mnohotvarý český. Ochr Přírody 1:18–23

Menges ES, Dolan RW (1998) Demographic viability of populations of *Silene regia* in midwestern prairies: relationships with fire management, genetic variation, geographic location, population size and isolation. J Ecol 86:63–78. https://doi.org/10.1046/j.1365-2745.1998.00234.x

Metzger JP et al (2009) Time-lag in biological responses to landscape changes in a highly dynamic Atlantic forest region. Biol Conserv 142:1166–1177. https://doi.org/10.1016/j.biocon.2009.01.033

Moeller DA, Geber MA (2005) Ecological context of the evolution of self-pollination in *Clarkia xantiana*: population size, plant communities, and reproductive assurance. Evolution 59:786–799. https://doi.org/10.1554/04-656

Münzbergová Z, Plačková I (2010) Seed mass and population characteristics interact to determine performance of *Scorzonera hispanica* under common garden conditions. Flora 205:552–559. https://doi.org/10.1016/j.flora.2010.04.001

Münzbergová Z, Cousins SAO, Herben T, Plačková I, Mildén M, Ehrlén J (2013) Historical habitat connectivity affects current genetic structure in a grassland species. Plant Biol 15:195–202. https://doi.org/10.1111/j.1438-8677.2012.00601.x

Nei M, Maruyama T, Chakraborty R (1975) Bottleneck effect and genetic variability in populations. Evolution 29:1–10. https://doi.org/10.2307/2407137

Oostermeijer J, Brugman M, DeBoer E, DenNijs H (1996) Temporal and spatial variation in the demography of *Gentiana pneumonanthe*, a rare perennial herb. J Ecol 84:153–166. https://doi.org/10.2307/2261351

Plue J, Vandepitte K, Honnay O, Cousins S (2017) Does the seed bank contribute to the build-up of a genetic extinction debt in the grassland perennial *Campanula rotundifolia*? Ann Bot 120:373–385. https://doi.org/10.1093/aob/mcx057

Prentice HC, Lonn M, Rosquist G, Ihse M, Kindstrom M (2006) Gene diversity in a fragmented population of *Briza media*: grassland continuity in a landscape context. J Ecol 94:87–97. https://doi.org/10.1111/j.1365-2745.2005.01054.x

Putz CM, Schmid C, Reisch C (2015) Living in isolation—population structure, reproduction, and genetic variation of the endangered plant species *Dianthus gratianopolitanus* (Cheddar pink). Ecol Evol 5:3610–3621. https://doi.org/10.1002/ece3.1611

R Development Core Team (2011) A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Radespiel U, Bruford MW (2014) Fragmentation genetics of rainforest animals: insights from recent studies. Conserv Genet 15:245–260. https://doi.org/10.1007/s10592-013-0550-3

Reed DH, Frankham R (2003) Correlation between fitness and genetic diversity. Conserv Biol 17:230–237. https://doi.org/10.1046/j.1523-1739.2003.01236.x

Reisch C et al (2017) Genetic diversity of calcareous grassland plant species depends on historical landscape configuration. BMC Ecol. https://doi.org/10.1186/s12898-017-0129-9

Reisch C, Schmid C, Hartig F (2018) A comparison of methods for estimating plant population size. Biodivers Conserv. https://doi.org/10.1007/s10531-018-1522-1

Reitschläger J (2000) Ekologické vazby a opylovací strategie ohroženého druhu *Gentianella bohemica* (hořeček český). Master thesis, Biological Faculty, University of South Bohemia, Czech Republic

Rosengren F, Cronberg N, Reitalu T, Prentice HC (2013) Genetic variation in the moss *Homalothecium lutescens* in relation to habitat age and structure. Botany 91:431–441. https://doi.org/10.1139/cjb-2012-0258

Ruane LG, Rotzin AT, Congleton PH (2014) Floral display size, conspecific density and florivory affect fruit set in natural populations of *Phlox hirsuta*, an endangered species. Ann Bot 113:887–893. https://doi.org/10.1093/aob/mcu007

Ryttäri T, Kukk Ü, Kull T, Jäkäläniemi A, Reitalu M (2003) Monitoring of threatened vascular plants in Estonia and Finland—methods and experiences. https://helda.helsinki.fi/bitstream/handle/10138/40538/FE_659.pdf?sequence=1

Šmídová A, Münzbergová Z, Plačková I (2011) Genetic diversity of a relict plant species, *Ligularia sibirica* (L.) Cass. (Asteraceae). Flora 206:151–157. https://doi.org/10.1016/j.flora.2010.03.003

Sork VL, Smouse PE (2006) Genetic analysis of landscape connectivity in tree populations. Landsc Ecol 21:821–836. https://doi.org/10.1007/s10980-005-5415-9

Šurinová M, Brabec J, Münzbergová Z (2017) Development of SSR markers by 454 sequencing in the endemic species *Gentianella praecox* subsp. *bohemica* (Gentianaceae). Appl Plant Sci 5:1600114. https://doi.org/10.3732/apps.1600114

Takkis K, Partel M, Saar L, Helm A (2013) Extinction debt in a common grassland species: immediate and delayed responses of plant and population fitness. Plant Ecol 214:953–963. https://doi.org/10.1007/s11258-013-0221-y

ter Braak C, Šmilauer P (1998) Canoco reference manual and users guide to Canoco for Windows: Software for canonical community ordination (version 4). Microcomputer Power, Ithaca

Tison JL, Blennow V, Palkopoulou E, Gustafsson P, Roos A, Dalen L (2015) Population structure and recent temporal changes in genetic variation in Eurasian otters from Sweden. Conserv Genet 16:371–384. https://doi.org/10.1007/s10592-014-0664-2

Turoňová D (2015) Metodika monitoringu a mapování kriticky ohrožených druhů cévnatých rostlin. http://www.biomonitoring.cz/?strankaID=6966

van der Meer S, Dahlgren JP, Mildén M, Ehrlén J (2014) Differential effects of abandonment on the demography of the grassland perennial *Succisa pratensis*. Popul Ecol 56:151–160. https://doi.org/10.1007/s10144-013-0400-7

Van Geert A, Van Rossum F, Triest L (2008) Genetic diversity in adult and seedling populations of *Primula vulgaris* in a fragmented agricultural landscape. Conserv Genet 9:845–853. https://doi.org/10.1007/s10592-007-9409-9

Wolf AT, Harrison SP (2001) Effects of habitat size and patch isolation on reproductive success of the serpentine morning glory. Conserv Biol 15:111–121. https://doi.org/10.1046/j.1523-1739.2001.99341.x

Young AG, Brown AHD, Zich FA (1999) Genetic structure of fragmented populations of the endangered daisy *Rutidosis leptorrhynchoides*. Conserv Biol 13:256–265. https://doi.org/10.1046/j.1523-1739.1999.013002256.x

Zenboudji S et al (2016) Conservation of the endangered Mediterranean tortoise *Testudo hermanni hermanni*: the contribution of population genetics and historical demography. Biol Cons 195:279–291. https://doi.org/10.1016/j.biocon.2016.01.007

Study 7: Low genetic differentiation despite high fragmentation in the endemic serpentinophyte *Minuartia smejkalii* (*M. verna* agg. Caryophyllaceae) revealed by RADSeq SNP markers

Conservation Genetics (2020) 21:187–198
https://doi.org/10.1007/s10592-019-01239-4

RESEARCH ARTICLE

Low genetic differentiation despite high fragmentation in the endemic serpentinophyte *Minuartia smejkalii* (*M. verna* agg., Caryophyllaceae) revealed by RADSeq SNP markers

Bojana Stojanova[1] · Maria Šurinová[2,3] · Vojtěch Zeisek[2,3] · Zuzana Münzbergová[2,3] · Hana Pánková[2]

Received: 28 May 2019 / Accepted: 20 November 2019 / Published online: 2 January 2020
© Springer Nature B.V. 2020

Abstract
Minuartia smejkalii is an obligate serpentinophyte plant endemic to the Czech Republic. Since the 1960s, the species' habitat has undergone strong human-mediated fragmentation, resulting in extinction of some populations and dramatic size reduction of the remaining populations. Thus, contrary to the typically stable serpentine habitats, *M. smejkalii* habitats underwent a recent and severe decline, which can exacerbate the effects of fragmentation on population genetic structure. We examined the genetic structure of all known *M. smejkalii* populations and two populations of *M. corcontica* and *M. caespitosa*, which are closely related, using RADSeq. The results indicate low, but clear differentiation among the three species, thus supporting the status of *M. smejkalii* as an independent taxon, though more extensive analysis of the whole group is needed. We further show high genetic diversity within *M. smejkalii* populations, low to moderate among-populations differentiation, and moderate regional differentiation. This could be due to the outcrossing mating system of *M. smejkalii* promoting high levels of gene flow and historical factors (multiple founder events, a recent bottleneck and/or a genetic time lag). We finally demonstrate that 2–3% of the markers show differentiation patterns consistent with divergent selection, suggesting that some local adaptation might have occured in *M. smejkalii*. Based on our observations, but without any experimental testing for local adaptation, if a conservation action is to be carried out, we recommend strictly separating the material from the two regions, and if possible, separating the populations within a region.

Keywords Caryophyllaceae · SNP markers · Anonymous loci · Relict species · Edaphic isolation · Species conservation

Introduction

Securing species persistence under the threat of human-mediated habitat fragmentation and reduction is a major concern for conservation biologists (Fahrig 2003; Lindenmayer and Fischer 2013). Classical in situ conservation approaches consist of restoration of habitat and species diversity, maintaining or increasing population sizes, and reintroducing species in locations where they went extinct. Lately, conservation policies are increasingly focusing on the preservation of genetic diversity in threatened species (Rodríguez-Quilón et al. 2016). Because it reduces population size and connectivity, habitat fragmentation can erode genetic variation through increased genetic drift and inbreeding, and reduced gene flow (Young et al. 1996). In endangered species in particular, high genetic diversity is important to buffer the potential negative impacts of environmental change, increase the adaptive potential of populations, and reduce the risk of inbreeding depression (Mimura et al. 2017).

Endemic species are a priority category for conservation, since their restricted geographic distribution makes them particularly vulnerable to habitat fragmentation and reduction (Wolf 2001; Wolf and Harrison 2001). Maintaining high genetic diversity is crucial for the long-term conservation of endemic species, as they rarely, if at all, migrate towards new

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10592-019-01239-4) contains supplementary material, which is available to authorized users.

✉ Bojana Stojanova
bojana.stojanova@gmail.com

[1] Department of Biology and Ecology, Faculty of Science, University of Ostrava, Ostrava, Czech Republic

[2] Institute of Botany, Academy of Sciences of the Czech Republic, Průhonice, Czech Republic

[3] Department of Botany, Faculty of Science, Charles University, Prague, Czech Republic

habitats when facing habitat degradation. A typical example of such a scenario is obligate serpentinophyte species, whose survival is bound to serpentine soils in nature. Serpentine soils have an ubiquitous but highly patchy distribution, covering less than 1% of the total dryland, which limits the possibility for migration of serpentine-adapted populations. These soils are characterized by high concentrations of toxic heavy metals (magnesium, nickel, cobalt, chromium), low levels of nutrients (potassium, phosphorous, nitrogen), and are often steep, rocky and subject to erosion. These conditions are frequently linked with low water retention. Thus, populations growing on serpentine soils experience strong selective pressures, as they need to adapt to the soil chemical composition, drought, and high temperatures (Linhart and Grant 1996; Brady et al. 2005). Due to the extreme conditions, the vegetation on serpentine soils is usually sparse, which results in a less competitive environment. As a consequence, serpentine-adapted populations, even when capable of growing on non-serpentine soils alone, are quickly extirpated because they cannot outcompete the non-serpentine species, likely because of a trade-off between stress tolerance and competitive abilities (Kruckeberg 1984; Harrison and Rajakaruna 2011).

The strong selective pressure and spatial isolation of serpentine soils are expected to result in low genetic diversity in serpentine species through bottlenecks, founder effects, or rapid elimination of poorly adapted genotypes (Whittaker 1954; Moore et al. 2013). Yet a remarkable number of serpentine species show high genetic variation within populations and small regional scales that are equal to or surpass genetic variation of their non-serpentine congeners (Ducousso et al. 1990; Linhart and Grant 1996; Mengoni et al. 2000; Mattner et al. 2002; Quintela-Sabarís et al. 2010; Moore et al. 2013). Possible explanations for this are gene flow from neighboring serpentine and non-serpentine habitats (if the species is not a strict serpentinophyte), multiple events of colonization in the past when serpentine soils acted as refugia for non-competitive genotypes, or the accumulation of new mutations and variants if the serpentine populations were isolated for a long time (Ducousso et al. 1990; Mengoni et al. 2000; Słomka et al. 2011). Typical serpentine habitats, although scattered, are usually stable over long time periods (Cheptou et al. 2017), which can help to maintain and build up genetic diversity of serpentine species.

A disjunct geographic range encompassing a wide edaphic niche ranging from calcareous and siliceous substrates to serpentines soils is well exemplified in the *Minuartia verna* agg. The *M. verna* agg. includes perennial taxa widely distributed in the Northern hemisphere (Meusel et al. 1965), but confined to various open habitats. The complex was widely distributed during the last glacial maxima, but in the Holocene its range retracted to so-called interglacial refugia on various extreme substrates, including serpentine soils, by the changing climatic conditions and increased competition from other species (Novák 1960). Being subjected to long-term geographic isolation and highly stressful serpentine soil conditions, the relict populations of *M. verna* agg. became genetically isolated through genetic drift, mutation and selection, resulting in several relict species (sensu biogeographic relicts, Lomolino et al. 2017) within the aggregate (Novák 1960; Mayer et al. 1994; Chytrý et al. 2017). Depending on the taxonomic concept, the *M. verna* agg. complex consists of between 3 and 15 species, often with contrasting edaphic preferences and geographic ranges. One of these is species is *Minuarita smejkalii*, an endemic to the Czech Republic and confined to small serpentine outcrops in highlands of the Central part of the Czech Republic (Dvořáková 1988). Two other species of the *M. verna* agg. complex are present in the Czech Republic: *M. corcontica* is an endemic occurring exclusively in the in Krkonoše Mountains (Dvořáková 2003), and *M. caespitosa* has a distribution which extends to Germany and Belgium (Kubát 2002). Up until now, the distinction among the three *Minuartia* species has only been made based on morphology (Dvořáková 1988, 2003), and no genetic comparisons have yet been made.

Being an obligate serpentinophyte, *M. smejkalii*'s natural occurrences are highly restricted by the patchy distribution of serpentine soils. Known habitats of *M. smejkalii* have been dramatically reduced in size and fragmented since the 1960–1970 by human actions. Thus, contrary to their typically fragmented but stable-over-time serpentine habitats, the habitat of *M. smejkalii* underwent a recent destruction followed by strong population decline. This could result in a severe reduction of genetic diversity through genetic drift, inbreeding, new and strong selective pressures, possibly leading to an extinction vortex (Cheptou et al. 2017).

In this study, we used codominant RADseq SNP markers to genotype individuals from all known localities of *M. smejkalii*, as well as two Czech populations of each *M. caespitosa* and *M. corcontica*. Our first aim was to disentangle the three congeneric species occurring within the Czech Republic based on the SNP markers, in order to contribute to the understanding of the taxonomic status of *M. smejkalii* and thus help conservation by clarifying its IUCN status. We then focused on the structuring of genetic diversity within the species of *M. smejkalii* and examined how it is partitioned on different geographic scales—between regions, among populations within a region, and among subpopulations. We also tested for the presence of outlier loci in our data set, which are putatively under selection, or at the very least population- or region-specific (private) alleles (Foll and Gaggiotti 2008). Altogether, these results are the first step into developing an efficient conservation strategy for *M. smejkalii* that maximizes its genetic diversity while avoiding deleterious effects of inbreeding or outbreeding depression.

Materials and methods

Population description and sampling

Minuartia smejkalii (Dvořáková) is a Czech-endemic species, with a self-compatible, but predominantly outcrossing mating system. *M. smejkalii* belongs to the species of priority European interest according to Habitats directive 92/43/EEC, Annex II. It is also included in the Convention on the Conservation of European Wildlife and Natural Habitats and the IUCN international Red list (Bilz 2011). According to the Czech laws, *M. smejkalii* is critically endangered (Grulich 2012). The distribution of *M. smejkalii* has historically been limited to three regions, all occurring within an area of 500 km^2: Borecká skalka Natural Monument, Hadce u Želivky National Natural Monument and Hadce u Hrnčíř Natural Monument (the latter two hereafter referred to as Želivka and Hrnčíře). Populations in the Borecká skalka region went extinct because of mining activities and the expansion of self-seeded trees in the 1960s. More than half of the available serpentine rocks suitable for the species in the Želivka region were destroyed by the construction of a motorway and a dam between 1965 and 1975 (Pešout 2001). In Hrnčíře, *M. smejkalii* occurred on several serpentine localities but they were destroyed by agricultural intensification in the 1960s. Currently, *M. smejkalii* in Hrnčíře survives only on one site which is an open area in a cultural forest.

In the Želivka region *M. smejkalii* occurs on rock platforms, crevices, and shallow serpentine soils, and fragmented grassland patches in open pine forests with sporadic vegetation (*Asplenion serpentinii* and *A. cuneifolii-Armerion serpentini* communities). There are six geographically distinct populations in the Želivka region: Z1–5 which are all within 600 m from of each other and Z6, which is located about 2000 m to the east. Populations Z2, Z3 and Z4 were further divided in subpopulations, which correspond to distinctive rock platforms (Table 1; Fig. 1). The total population size in the Želivka region varied between 400 and 1200 in the last 13 years (Pánková, pers.obs). Populations on the border of the region are the smallest or have already gone extinct. In the Hrnčíře region, *M. smejkalii* grows in one population (marked as H) divided in two subpopulations—one in cultural pine forest openings with 21–97 individuals in the last 13 years, and a new subpopulation, naturally established in 2011 along a nearby forest path after tree cutting with 110–227 individuals (Table 1; Fig. 1).

Two other *Minuartia* species from the *M. verna* agg. complex occurring in the Czech Republic were also included in the study. *Minuartia caespitosa* has a broad distribution from eastern Belgium, through north-western Germany to northern Czech Republic (Hejný and Slavík 2003). *Minuartia caespitosa* occurs in the Czech Republic in a 2 km^2 area in Hradčanské stěny (Hradčany further in the text), where they occur in calcareous grasslands on small, unstable sand dunes with sporadic vegetation at the bottom of sandstone rocks (Sádlo et al. 2011). For the purpose of this study, two populations 0.4 km from each other in the region of Hradčany were sampled. *Minuartia corcontica* is endemic to the Czech Republic that at present grows in two populations in the Krkonoše National Park (Krkonoše further in the text), on pyroxen gneiss with a high mineral and sulphide content covering an area of approximately 0.06 km^2. Only one population in this area (divided in two subpopulations), is accessible for sampling and was thus sampled for this study (Fig. 1).

Table 1 Population genetics characteristics

Species	Region	Subpops	Pop	N	λ	Ia	% nonHW	H_{obs}	H_{exp}	F_{IS}
M. corcontica	Krkonoše	1	MC	5	0.8	5	5.687	0.206	0.154	−0.335
		1	VC	5	0.8	4.07	6.299	0.221	0.173	−0.281
M. caespitosa	Hradčany	1	SK1	5	0.8	4.63	5.074	0.214	0.174	−0.232
		1	SK2	4	0.75	2.85	5.774	0.213	0.174	−0.224
M. smejkalii	Želivka	1	Z1	9	0.889	1.93	7.962	0.210	0.192	−0.095
		11	Z2	28	0.964	2.5	5.424	0.259	0.217	−0.193
		1	Z3	9	0.889	1.47	20.385	0.269	0.206	−0.305
		2	Z4	10	0.9	2.16	9.099	0.247	0.204	−0.213
		1	Z5	10	0.9	2.92	9.274	0.244	0.213	−0.146
		1	Z6	9	0.889	2.47	7.612	0.213	0.208	−0.027
	Hrnčíře	1	H	14	0.929	10.88	12.073	0.255	0.210	−0.218
			Total	108	0.991					

N number of genotyped individuals, *λ* Simson index of the probability that two random genotypes in the population are different, *Ia* index of association, *%nonHW* percentage of the loci in the population that are not in Hardy–Weinberg equilibrium, H_{obs} observed heterozygosity, H_{exp} expected heterozygosity, F_{IS} fixation index

Fig. 1 Map of sampled populations and regions. **a** Distribution of the species of the *Minuartia verna* agg. complex in the Czech Republic. **b** Distribution of the populations of *M. smejkalii* in the Želivka region

For genotyping we sampled vegetative tissue that consisted of a stem fragment with a few green leaves. The number of individuals sampled per population varied according to the species and population size and structure. In populations Z1, Z5 and Z6, for which a subpopulation structure was not defined, up to ten individuals per population were sampled. In populations that were further divided in subpopulations, three (Z2 with 11 subpopulations) to five (H, Z3, and Z4 with 2 subpopulations) individuals were sampled per subpopulation. For *M. corcontica,* five individuals per subpopulation, i.e. 10 samples in total were collected (Table 1). The number of samples was limited by the low population size and difficulty in accessing the species. We sampled the same number of plants for *M. caespitosa*, i.e. five individuals per population. While this species is more widespread, for the purpose of this study we only sampled the population that is geographically the closest to *M. smejkalii*. Altogether, this resulted in 119 *Minuartia* samples.

DNA extraction and RADSeq genotyping

Total genomic DNA of 119 individuals was extracted from dehydrated leaves using a modified cetyltrimethylammonium bromide (CTAB) method (Lodhi et al. 1994). Library preparation and sequencing were made by SNPsaurus, LLC, following a previously established protocol (Russello et al. 2015). Genomic DNA was converted into nextRAD genotyping-by-sequencing libraries. Genomic DNA was first fragmented with Nextera reagent (Illumina, Inc.), which also ligates short adapter sequences to the ends of the fragments. The Nextera reaction was scaled for fragmenting 10 ng of genomic DNA, although 20 ng of genomic DNA was used for input to compensate for the amount of degraded DNA in the samples and to increase fragment sizes. Fragmented DNA was then amplified for 26 cycles at 73°, with one of the primers matching the adapter and extending nine nucleotides into the genomic DNA with the selective sequence GTGTAG AGC. Thus, only fragments starting with a sequence that can be hybridized by the selective sequence of the primer were efficiently amplified. The indexes used and the PCR protocol are part of Nextera DNA library preparation kit (Illumina, Inc.). No technical replicates were included. The nextRAD libraries were sequenced on a HiSeq4000 with two lanes (60 and 59 individuals in each line). The nextRAD libraries were sequenced on a HiSeq4000 with two lanes of 150 bp reads (University of Oregon). The genotyping analysis used custom scripts (SNPsaurus, LLC) that trimmed the reads using bbduk (Bushnell 2014).

A de novo reference genome was created by collecting 10 million reads in total, evenly from the samples, and excluding loci that had a depth of coverage of fewer than five (insufficient coverage) or more than 1000 (derived from clonal amplification or repetitive elements; Morris et al. 2016). The remaining loci were then aligned to each other to identify allelic loci and collapse allelic haplotypes to a single locus. All reads were mapped to the reference with an alignment identity threshold of 90% using bbmap (Bushnell 2014). Genotype calling was done using Samtools and bcftools. The vcf was filtered to remove alleles with a population frequency of $< 3\%$ which are likely due to a genotyping error. Loci that were heterozygous in all samples or had more than 2 alleles per sample (suggesting collapsed paralogs) were removed. The absence of artifacts was confirmed by counting SNPs at each read nucleotide position and determining that SNP number did not increase with reduced base quality at the end of the read.

Statistical analyses

Because our RADSeq loci are anonymous, we tested for candidate loci under selection (outlier loci) using BayeScan 2.1 (Foll and Gaggiotti 2008). Twenty pilot runs of 5000 iterations were run, and the estimates were made from 5000 iterations with a thinning interval of 10, after a burn-in period of 10,000 iterations. We set a prior odds ratio of 10 (prior belief that a selection model is 1/10 as likely as a neutral model for a given SNP), which is considered as a strong evidence for selection (Foll and Gaggiotti 2008).

For the purposes of the evaluation of the genetic clustering of the individuals, we considered statistical techniques based on Bayesian clustering as very effective because they do not involve a priori hypotheses about sample clustering. We used Bayesian K-means clustering as described for Discriminant Analysis of Principal Components (DAPC, Jombart et al. 2010). It processes PC-scaled data by K-means clustering and maximizes the manifestation of the major pattern involved in the data by subsequent discriminant analysis (Jombart et al. 2010). The best clustering K was chosen according to step-wise decrease of BIC (the biggest decrease) with regard to absolute minimal value of the BIC, as discussed in Jombart et al. (2010).

Most computations were performed in R 3.4 (R Core Team 2014). We used packages ade4 (Dray and Dufour 2007), adegenet (Jombart 2008), APE (Paradis et al. 2004), pegas (Paradis 2010), Poppr (Kamvar et al. 2014) and vcfR (Knaus and Grünwald 2017). We calculated the distribution and diversity of multi locus genotypes (MLGs) within species (package poppr), F-statistics (package hierfstat), allelic richness (package poppr), index of association (Brown et al. 1980; Smith et al. 1993; Agapow and Burt 2001; package poppr), departure from Hardy–Weinberg equilibrium (HWE, Jombart 2008; package pegas), Principal Coordinate Analysis (PCoA, Dray et al. 2003; packages ade4 and adegenet), and K-means clustering (with 10,000,000 iterations and maximal $K = 35$, Jombart et al. 2010; package adegenet). We applied NeigborNet network analysis using Uncorrected_P

distance matrix in SplitsTree (Huson and Bryant 2005) to better display presumably tree-like relationships among samples.

Because F_{ST} estimates in small populations are dependent on the sample size, for the F_{ST} comparisons among populations of *M. smejkalii* we randomly sampled ten individuals in population Z2 which had 30 individuals in total. We did not estimate F_{ST} among the populations of different species, as they have been completely geographically isolated since the last glacial period, which effectively prevents gene flow among them.

Testing of how genetic diversity was structured between species, geographic regions, populations and subpopulations was conducted using analysis of molecular variance (AMOVA). Because the data set was not balanced, in particular with the number of subpopulations and populations differing among regions and species, we first examined how the variance was distributed in the entire data set (all populations), including all hierarchical levels, and then we ran a separate AMOVA for the *M. smejkalii* data only (populations Z1–Z6, H; hierarchical levels—region, population, subpopulation), for the Želivka populations only (populations Z1–Z6; hierarchical levels population, subpopulation), and for populations Z2 and Z3 each (subpopulation).

Finally, we tested for the occurrence of recent population bottlenecks using the software BOTTLENECK 1.2.02 (Piry et al. 1999), assuming an infinite allele model of mutation which is the most suitable for biallelic SNP markers (Cornuet and Luikart 1996; Kogura et al. 2011). Because sample sizes were smaller than the recommended sizes except for population Z2, we only analyzed population Z2 on its own, and we also analyzed all Želivka populations as a single population to increase the power of the analyses. To determine whether observed population heterozygosity was significantly greater than that expected under mutation-drift balance, we applied the sign test and the standardized difference test based on 10,000 iterations.

Results

SNP loci characteristics

Out of the 119 genotyped individuals, nine were eliminated because they had over 75% missing data—one from *M. caespitosa* (population SK2), two from Z2 and six from population H. There were 12,599 potential SNP sites (variants), distributed over 2800 short reads (contigs). The average depth of coverage was 67.15 (standard error 2.43, Online resource 3). For the subsequent analyses, only loci with < 2.5% missing data (null alleles) were retained, resulting in 1 143 polymorphic SNP sites (loci).

Identification of putatively adaptive loci

Out of the 1143 analyzed loci, a negligible number were under divergent selection among species and between the two regions of *M. smejkalii* (six and three respectively). We detected 20 loci among populations of *M. smejkalii* (1.7% of the total number of analyzed loci), and 34 among subpopulations (3.2%, Online resource 2) under putative divergent selection. We thus recalculated F_{ST} without outlier loci for populations and subpopulations only.

Population genetics and demography

Genotypic richness was at its maximal value in all populations since each individual had a unique multilocus genotype. The index of association was between 1.5 and 3 in the Želivka populations, and 10.8 in population H, suggesting higher linkage disequilibrium in the latter. In the other species, the index of association was between 3 and 4.5 for both *M. corcontica* and *M. caespitosa* (Table 1). Testing for Hardy–Weinberg equilibrium within each population showed that between 5 and 20% of the loci were not in equilibrium. This was likely due to heterozygote excess which was observed for all populations (negative F_{IS}) to different extents.

Assuming an infinite allele mutation model, the number of loci exhibiting heterozygosity excess relative to the expected heterozygosity under mutation-drift balance was significantly higher (p value < 0.001 for all statistical tests and all analyzed samples), suggesting the occurrence of a recent population bottleneck. Note that because of the low sample sizes for *M. corcontica*, *M. caespitosa* and *M. smejkalii* in the Hrnčíře region, testing for bottlenecks could only be done in the Želivka region and in population Z2.

Among-species differentiation

According to the K-means clustering, the most plausible number of clusters was two when all species are analyzed together (Online resource 3). With two clusters, the DAPC results showed a clear separation between *M. smejkalii* and the other species, with all specimens being correctly assigned to their designated species (Fig. 2). A biologically meaningful structure was also observed with four clusters: *M. concortica, M. caespitosa*, and the regions Želivka and Hrnčíře, although the clustering was not supported by the step-wise selection method of K. The SpitsTree clustered all individuals of *M. corcontica* and *M. caespitosa* in two sister clades separated from all *M. smejkalii* individuals (Fig. 3). The AMOVA results showed that most of the genetic variation was found within populations (individual level, Fig. 4). About 20% of the genetic variation was explained by the species, and population and subpopulation explained a small

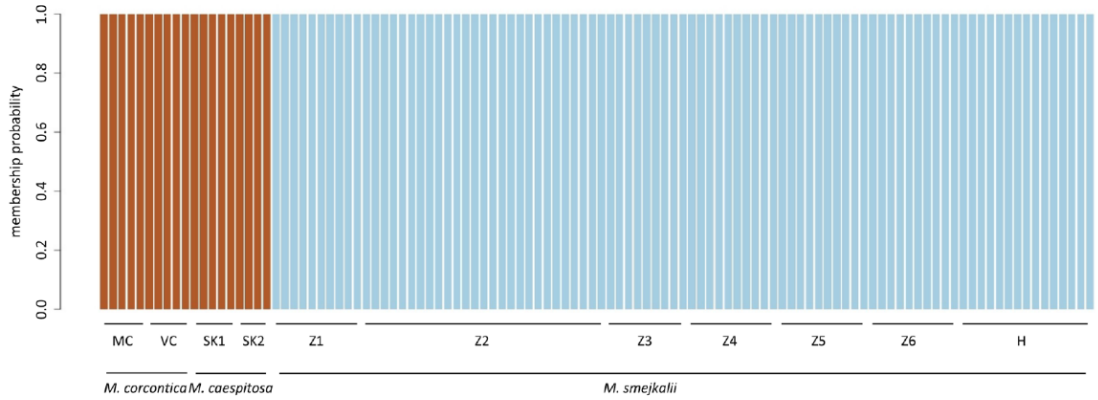


Fig. 2 Results of K-means clustering with two clusters. First line below the x-axis corresponds to population names, second to species

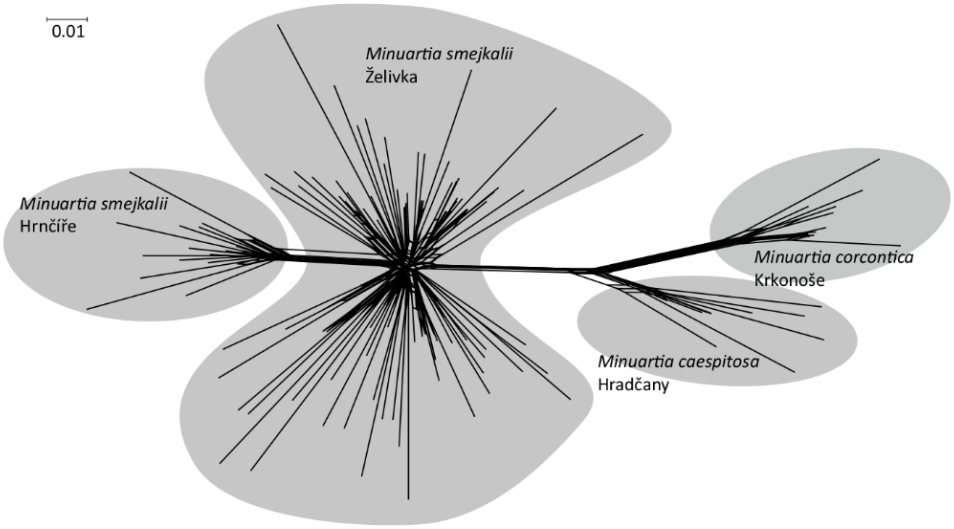


Fig. 3 SplitsTree network delimiting species (*M. corcontica*, *M. caespitosa* and *M. smejkalii*). Within populations of *M. smejkalii*, samples from Hrnčíře are nested within samples from Želivka. The population from Hrnčíře is strongly differentiated from the Želivka population, especially in comparison with the other species (*M. caespitosa*, *M. corcontica*)

(2.7 and 1.8%, respectively) but significant amount of the genetic variation.

Among-region and population differentiation

Global F_{ST} among populations of *M. smejkalii* was 0.036. Pairwise F_{ST} between populations were quite similar for all populations, ranging from 0.027 to 0.052, including population H which comes from a different region than the populations Z1–Z6. Removing the loci putatively under selection from the estimates resulted in lower global F_{ST} (0.022) and overall lower pairwise F_{ST} among populations (Table 2).

According to the K-means clustering, the best number of clusters for *M. smejkalii* populations were analyzed was one—the BIC for K = 2 was not supported by the step-wise selection for K (Online resource 3). The SplitsTree clustered individuals of the Hrnčíře population in one clade, nested in the Želivka clade (Fig. 3). The AMOVA results showed that almost 77% of the total genetic variation was found within subpopulations, 18% was between regions, 1.9% among

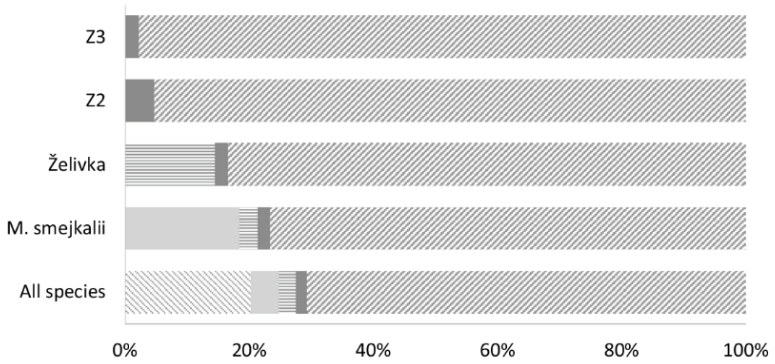


Fig. 4 Percentage of variation explained by the embedded hierarchical levels (species, region, population, subpopulation) in different AMOVA. All species—analysis of the entire data set, *M. smejkalii*—only *M. smejkalii* populations, Želivka—only the six populations from Želivka, Z2 and Z3—only the populations divided into subpopulations (11 and two subpopulations of populations Z2 and Z3 respectively). Significant variance components are marked with an asterisk

	Species	Region	Population	Subpopulation	Individual
Z3	-	-	-	2.2	97.3
Z2	-	-	-	4.7*	95.3
Zelivka	-	-	14.5*	2.1*	83.4
M.smejkalii	-	18.4*	3.0	1.9*	76.7
All species	2.3*	4.5	2.7*	1.8*	70.7

Table 2 Pairwise population F_{ST} estimates for all loci (above the diagonal) and for putatively neutral loci only (below diagonal)

	Z1	Z2	Z3	Z4	Z5	Z6	H
Z1		0.039	0.043	0.043	0.048	0.038	0.052
Z2	0.035		0.027	0.028	0.035	0.033	0.037
Z3	0.040	0.027		0.035	0.038	0.039	0.040
Z4	0.036	0.023	0.031		0.041	0.036	0.038
Z5	0.038	0.026	0.030	0.033		0.034	0.047
Z6	0.034	0.026	0.036	0.030	0.030		0.045
H	0.038	0.028	0.029	0.030	0.034	0.032	

To avoid comparing samples of different size, in population Z2 ten individuals (out of 28) were randomly drawn and used for the estimates

subpopulations, whereas the population did not have significant effect on genetic diversity partitioning (Fig. 4).

Within region differentiation (Želivka)

Within the region of Želivka, the K-means clustering suggested that the best estimate of structure was obtained for K = 1, and the SplitsTree analysis did not detect any clustering by population. According to the AMOVA results, population and subpopulation were significant and explained 14.5% and 2.1% of the total variation, respectively (Fig. 4).

Discussion

Patterns of among-species genetic variation

Present-day conservation policies require a reliable taxonomic identification of the endangered taxa, especially for establishing conservation priorities when available resources are limited (McNeely 2002; Wege et al. 2015). The taxonomic status of the Czech endemic species *M. smejkalii* is not fully resolved, with a recent taxonomic

study classifying it as a synonym of the widespread *M. verna* ssp. *gerardii* (Dillenberger and Kadereit 2014). However, Dillenberger and Kadereit (2014) did not mention having any sample from the Czech Republic for their analysis and likely thus did not have *M. smejkalii* in their material at all. The best way for proper distinction between *Minuartia* species is the morphology of seeds (Hejný and Slavík 2003). Comparison of the seed morphology of *M. verna* ssp. *verna* (Germany 300–400 m above sea level), *M. verna* ssp. *gerardii* (Austrian Alps, more than 1200 m above sea level), and *M. smejkalii*, *M. caespitosa,* and *M. corcontica* from this study under electron microscope showed that each subspecies/species had different seed morphology, especially in its protuberance (Pánková and Machač unpubl).

Our clustering analyses clearly separated all *M. smejkalii* individuals from *M. corcontica* and *M. caespitosa*, the other two species which originated from *M. verna* agg. and occur in the Czech Republic, suggesting an independent taxonomic status of *M. smejkalii*. The exact taxonomic level of *M. smejkalii* (species, subspecies, ecotype) needs yet to be determined in a global treatment of the whole *M. verna* agg., which is the goal of an ongoing study which uses very broad taxonomic sampling of *Minuartia* in central Europe (Jindřich Chrtek, pers. com.). Although this expanded sampling might produce results which do not warrant its endemic status in the Czech Republic, the differentiation of *M. smejkalii* from the other *Minuartia* species of the *M. verna* agg. in the Czech Republic supports its treatment as a separate entity in conservation programs (Montalvo and Ellstrand 2000; Hufford and Mazer 2003; McKay et al. 2005).

Patterns of within species genetic variation

Minuartia smejkalii is an endemic taxon with fragmented distribution, reduced number of populations, small population size, growing in an unfavorable, serpentine soil environment, which can result in high genetic drift and inbreeding, and strong local selective pressure. This should cause reduced within population genetic diversity, and possibly increased among-population genetic variation (Hamrick and Godt 1989; Ellstrand and Elam 1993; Linhart and Grant 1996; Whitlock et al. 2016). Our AMOVA results showed however moderate regional differentiation, low to moderate population differentiation within the Želivka region, and high genetic variation within populations. Similar results have also been observed for other rare, endangered, and endemic species (Ellstrand and Elam 1993). For instance, *Petunia secreta*, a Brazilian endemic with a very similar distribution to *M. smejkalii*, showed genetic differentiation between two regions about 20 km apart, but not within regions, where the populations were < 0.2 km apart (Turchetto et al. 2016).

Genetic differentiation at relatively small geographic scales, such as those between the regions Želivka and Hrnčíře in this study, is commonly observed in serpentine species (Linhart and Grant 1996; Mattner et al. 2002). This regional differentiation could be caused by the lack of connectivity among serpentine patches between regions resulting in isolation by distance, and/or because of differences in regional environmental conditions, resulting in contrasting selective pressures. The levels of regional differentiation in *M. smejkalii* according to F_{ST} estimates in this study were between 0.028 and 0.052 (depending on the populations and loci included in the estimates), which is much lower than those found for other serpentine *Minuartia* species: 0.186 for *M. laricifolia* ssp *laricifolia*, 0.1–0.279 for *M. laricifolia* ssp *ophiolitica*, 0.249 for *M. biflora* and 0.895 for the selfing *M. rubella* (Borgen 1999; Moore et al. 2013). This could suggest some gene flow between regions in the past despite the ridge that separates the Želivka and Hrnčíře sites. The different F_{ST} estimates can also be due to the different types of markers used in different studies with contrasting polymorphism, genome coverage or dominance levels (AFLP, Moore et al. 2013; allozymes, Borgen 1999). However, F_{ST} estimates are also influenced by the sampling method, population size, history, or mating system, which makes it difficult to conclude on the extent to which the marker type influence the estimates (Mariette et al. 2002; Leipold et al. 2018).

Low levels of regional genetic differentiation however are not necessarily excluding the possibility of local adaptation. Mosca et al. (2012) showed that some conifer populations differentiated by < 5% according to molecular markers, could still show strong patterns of local adaptation in quantitative genetic traits. In line with this, we detected between 2 and 3% (depending on the levels of analysis) of outlier loci putatively under divergent selection suggesting the occurrence of local adaptation on the population level (nested within region). Interpreting outlier loci as adaptive should however be made with caution, as outliers can result also from low recombination rates due to non-evolutionary processes (Shafer et al. 2015). To confirm their adaptive value, the outlier loci need to be either correlated with fitness related traits, or their function needs to be identified using a reference genome (Fuentes-Pardo and Ruzzante 2017) which is not yet available for *M. smejkalii* or any other *Minuartia* species.

Genetic differentiation among the populations in the Želivka region was very low when all *M. smejkalii* species were analyzed with AMOVA, and moderate when only populations from Želivka were analyzed. Therefore, most of the genetic variation in our system, regardless of the AMOVA hierarchy was within populations. Similar results have also been observed for the Alpine populations of *M. laricifolia* spp. *laricifolia* (Moore et al. 2013). Several serpentine soil

specialist species also showed relatively low among-population genetic differentiation and high within-population genetic variation (Mengoni et al. 2000; Mattner et al. 2002; Quintela-Sabarís et al. 2010; Babst-Kostecka et al. 2014). More generally, population genetic differentiation of rare species is sometimes comparable and even greater than that of congeneric widespread species (Gitzendanner and Soltis 2000; Zawko et al. 2001; Ellis et al. 2006; Turchetto et al. 2016; Lanes et al. 2018) especially in perennial plants (Cabrera-Toledo et al. 2008).

High within- and low among-population genetic diversity can be a consequence of an outcrossing mating system and long life cycle (Nybom and Bartish 2000). Outcrossing species have in general F_{ST} values of 0.2 or less, (Mattner et al. 2002) and high within population genetic variation, in particular in dioecious or self-incompatible species (Cabrera-Toledo et al. 2008). Although *M. smejkalii* is self-compatible (Stojanova, pers.obs.), it does not have any mechanisms facilitating self-pollination: flowers isolated from pollinators tend to produce smaller seed sets than open pollinated flowers (Pánková, pers. obs.). Furthermore, some plants grown in a greenhouse show protogynous flowering patterns—early season flowers are mostly female, and hermaphrodite flowers develop later during the season (Stojanova, pers. obs.). Altogether, these results suggest a highly outcrossing mating system, which could explain the observed levels of genetic variation.

Another possible explanation for the observed patterns of genetic diversity in *M. smejkalii* is the effect of historical evolutionary processes. For instance, independent colonization of different sites, with multiple founder events or gene flow between serpentine and (now extinct) non-serpentine populations may have contributed to historically high levels of genetic variation in some serpentine species (Moore et al. 2013). This scenario seems unlikely for *M. smejkalii,* as no other related species (i.e. descendants from *M. verna* agg.) has been recorded in the region. High levels of genetic variation can also be maintained by genetic time lags, especially in long-lived species that only recently encountered habitat fragmentation (Cabrera-Toledo et al. 2008; Epps and Keyghobadi 2015). The habitat fragmentation of *M. smejkalii* began in the 1960, and the species could have had as little as four generations since then. Indeed, plants recorded in 2001 are still alive and reproducing at the time of this study, 18 years later. In line with this, the observed heterozygosity was higher than the expected heterozygosity in all populations, consistent with occurrence of a recent bottleneck. This suggests that the populations of *M. smejkalii* have still not reached a genetic equilibrium after the disturbance of their habitat, and the high levels of genetic diversity are likely a transient state (Cornuet and Luikart 1996). Thus, without conservation efforts the populations genetic diversity might decline in the future.

Conservation recommendations

Despite recent habitat fragmentation and population size reduction, *M. smejkalii* still maintains high levels of genetic diversity within populations, moderate regional differentiation and low to moderate among-populations differentiation. Determining the exact causes for the observed genetic structure requires further observations and experimental work to characterize the breeding system of the species in more detail, experimentally test for local adaptations, and detect potential effects or absence of selective pressure. It is, however, plausible that the high levels of genetic diversity are only a transient state following a recent population disturbance, and will not be maintained indefinitely if no conservation actions are undertaken.

The ex situ conservation efforts of *M. smejkalii* should focus on maintaining the observed genetic diversity while increasing population size and territory. This requires choosing appropriate seed source to enhance population size for each region. Based solely on the results of this study, it is difficult to unequivocally recommend a single conservation strategy.

Given the genetic differentiation between regions, we do not recommend mixing individuals from the regions of Želivka and Hrnčíře. Within a region, low among-population differentiation is taken as a justification for the use of transplants regardless of their population of origin, since they will not erode the local genetic structure of the extant populations (Cabrera-Toledo et al. 2008). However, patterns of molecular diversity are not necessarily related to adaptive traits (Reed and Frankham 2001; Kramer and Havens 2009). In line with this, we detected several outlier loci which could be compatible with patterns of divergent selection. Based on these observations, one might recommend a conservation strategy using exclusively material coming from the same population. A counter argument to this is the mating system of *M. smejkalii*, which is likely outcrossing. If this is indeed true, using only within-population/within-region transplants could increase the risk of inbreeding depression through biparental inbreeding. Experimental assays are currently being conducted to test for the existence of inbreeding and outbreeding depression in *M. smejkalii*, which will provide the necessary input for a reliable and efficient conservation strategy. In the meanwhile, if any conservation action is to be carried out, we recommend the use of individuals from the same region chosen to maximize the genetic diversity within populations.

Acknowledgements The study was supported by Project LIFE for Minuartia (Grant No. LIFE15NAT/CZ/000818), which is implemented with the financial contribution of the European Union, the Life program and with the financial contribution of the Ministry of Environment of the Czech Republic and party by and the long-term research development Project RVO 67985939.

Data availability Genomic data is available at BioSample, ID: SUB5620362.

Compliance with ethical standards

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

References

Agapow P-M, Burt A (2001) Indices of multilocus linkage disequilibrium. Mol Ecol Notes 1:101–102

Babst-Kostecka AA, Parisod C, Godé C, Vollenweider P, Pauwels M (2014) Patterns of genetic divergence among populations of the pseudometallophyte *Biscutella laevigata* from southern Poland. Plant Soil 383:245–256

Bilz M (2011) The IUCN Red List of Threatened Species. *Minuartia smejkalii. The IUCN Red List of Threatened Species 2011*

Borgen L (1999) Genetic variation in *Minuartia* (Caryophyllaceae) in Svalbard. Nord J Bot 19:179–192

Brady KU, Kruckeberg AR, Bradshaw HD Jr (2005) Evolutionary ecology of plant adaptation to serpentine soils. Annu Rev Ecol Evol Syst 36:243–266

Brown AHD, Feldman MW, Nevo E (1980) Multilocus structure of natural populations of *Hordeum spontaneum*. Genetics 96:523–536

Bushnell B (2014) BBMap: a fast, accurate, splice-aware aligner

Cabrera-Toledo D, Gonzalez-Astorga J, Vovides AP (2008) Heterozygote excess in ancient populations of the critically endangered *Dioon caputoi* (Zamiaceae, Cycadales) from central Mexico. Bot J Linn Soc 158:436–447

Cheptou P-O, Hargreaves AL, Bonte D, Jacquemyn H (2017) Adaptation to fragmentation: evolutionary dynamics driven by human influences. Philos Trans R Soc B 372:20160037

Chytrý M, Danihelka J, Kaplan Z, Pyšek P (2017) Flora and vegetation of the Czech Republic. Springer, Cham

Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics 144:2001–2014

Dillenberger MS, Kadereit JW (2014) Maximum polyphyly: multiple origins and delimitation with plesiomorphic characters require a new circumscription of Minuartia (Caryophyllaceae). Taxon 63:64–88

Dray S, Dufour A-B (2007) The ade4 package: implementing the duality diagram for ecologists. J Stat Softw 22:1–20

Dray S, Chessel D, Thioulouse J (2003) Co-inertia analysis and the linking of ecological data tables. Ecology 84:3078–3089

Ducousso A, Petit D, Valero M, Vernet P (1990) Genetic variation between and within populations of a perennial grass: *Arrhenatherum elatius*. Heredity 65:179

Dvořáková M (1988) Minuartia smejkalii, eine neue Art aus der *Minuartia gerardii-Gruppe* (Caryophyllaceae). Preslia 60:1–9

Dvořáková M (2003) *Minuartia pauciflora*, a Carpathian endemic species of the M-verna group. Preslia 75:349–356

Ellis JR, Pashley CH, Burke J, McCauley DE (2006) High genetic diversity in a rare and endangered sunflower as compared to a common congener. Mol Ecol 15:2345–2355

Ellstrand NC, Elam DR (1993) Population genetic consequences of small population size: implications for plant conservation. Annu Rev Ecol Syst 24:217–242

Epps CW, Keyghobadi N (2015) Landscape genetics in a changing world: disentangling historical and contemporary influences and inferring change. Mol Ecol 24:6021–6040

Fahrig L (2003) Effects of habitat fragmentation on biodiversity. Annu Rev Ecol Evol Syst 34:487–515

Foll M, Gaggiotti OE (2008) A genome scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics 180:977–993

Fuentes-Pardo AP, Ruzzante DE (2017) Whole-genome sequencing approaches for conservation biology: advantages, limitations and practical recommendations. Mol Ecol 26:5369–5406

Gitzendanner MA, Soltis PS (2000) Patterns of genetic variation in rare and widespread plant congeners. Am J Bot 87:783–792

Grulich V (2012) Red List of vascular plants of the Czech Republic. Preslia 84:631–645

Hamrick JL, Godt MJW (1989) Allozyme diversity in plant species. In: Brown AHD, Clegg MT, Kahler AL, Weir BS (eds) Plant population genetics, breeding and genetic resources. Sinauer Associates, Sunderland, pp 43–63

Harrison S, Rajakaruna N (2011) Serpentine: the evolution and ecology of a model system. University of California Press, Berkeley

Hejný S, Slavík B (2003) Květena ČR, díl 2. Academia, Praha

Hufford KM, Mazer SJ (2003) Plant ecotypes: genetic differentiation in the age of ecological restoration. Trends Ecol Evol 18:147–155

Huson DH, Bryant D (2005) Estimating phylogenetic trees and networks using SplitsTree 4. Manuscr. Prep. Softw. www.split.org. Accessed 22 Dec 2019

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet 11:94

Kamvar ZN, Tabima JF, Grünwald NJ (2014) Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2:e281

Knaus BJ, Grünwald NJ (2017) VCFR: a package to manipulate and visualize variant call format data in R. Mol Ecol Resour 17:44–53

Kogura Y, Seeb JE, Azuma N, Kudo H, Abe S, Kaeriyama M (2011) The genetic population structure of lacustrine sockeye salmon, *Oncorhynchus nerka*, in Japan as the endangered species. Environ Biol Fishes 92:539–550

Kramer AT, Havens K (2009) Plant conservation genetics in a changing world. Trends Plant Sci 14:599–607

Kruckeberg AR (1984) California serpentines: flora, vegetation, geology, soils, and management problems. University of California Press, Berkeley

Kubát K (2002) Klíč ke květeně České republiky. Acad, Praha

Lanes ÉC, Pope NS, Alves R, Carvalho Filho NM, Giannini TC, Giulietti AM, Imperatriz-Fonseca VL, Monteiro W, Oliveira G, Silva AR (2018) Landscape genomic conservation assessment of a narrow-endemic and a widespread morning glory from Amazonian savannas. Front Plant Sci 9:532

Leipold M, Tausch S, Hirtreiter M, Poschlod P, Reisch C (2018) Sampling for conservation genetics: how many loci and individuals are needed to determine the genetic diversity of plant populations using AFLP? Conserv Genet Resour. https://doi.org/10.1007/s12686-018-1069-1

Lindenmayer DB, Fischer J (2013) Habitat fragmentation and landscape change: an ecological and conservation synthesis. Island Press, Washington, DC

Linhart YB, Grant MC (1996) Evolutionary significance of local genetic differentiation in plants. Annu Rev Ecol Syst 27:237–277

Lodhi MA, Ye G-N, Weeden NF, Reisch BI (1994) A simple and efficient method for DNA extraction from grapevine cultivars and Vitis species. Plant Mol Biol Report 12:6–13

Lomolino MV, Riddle BR, Whittaker RJ (2017) Biogeography: biological diversity across space and time. Sinauer Associates Inc., Sunderland

Mariette S, Le Corre V, Austerlitz F, Kremer A (2002) Sampling within the genome for measuring within-population diversity: trade-offs between markers. Mol Ecol 11:1145–1156

Mattner J, Zawko G, Rossetto M, Krauss SL, Dixon KW, Sivasithamparam K (2002) Conservation genetics and implications for restoration of *Hemigenia exilis* (Lamiaceae), a serpentine endemic from Western Australia. Biol Conserv 107:37–45

Mayer MS, Soltis PS, Soltis DE (1994) The evolution of the *Streptanthus glandulosus* complex (Cruciferae): genetic divergence and gene flow in serpentine endemics. Am J Bot 81:1288–1299

McKay JK, Christian CE, Harrison S, Rice KJ (2005) "How local is local?"—a review of practical and conceptual issues in the genetics of restoration. Restor Ecol 13:432–440

McNeely JA (2002) The role of taxonomy in conserving biodiversity. J Nat Conserv 10:145–153

Mengoni A, Gonnelli C, Galardi F, Gabbrielli R, Bazzicalupo M (2000) Genetic diversity and heavy metal tolerance in populations of *Silene paradoxa* L. (Caryophyllaceae): a random amplified polymorphic DNA analysis. Mol Ecol 9:1319–1324

Meusel H, Jäger E, Rauschert J, Weinert E (1965) Vergleichende Chorologie der zentraleuropäischen Flora. Jena

Mimura M, Yahara T, Faith DP, Vázquez-Domínguez E, Colautti RI, Araki H, Javadi F, Núñez-Farfán J, Mori AS, Zhou S (2017) Understanding and monitoring the consequences of human impacts on intraspecific variation. Evol Appl 10:121–139

Montalvo AM, Ellstrand NC (2000) Transplantation of the Subshrub *Lotus scoparius*: testing the home-site advantage hypothesis. Conserv Biol 14:1034–1045

Moore AJ, Merges D, Kadereit JW (2013) The origin of the serpentine endemic *Minuartia laricifolia* subsp. *ophiolitica* by vicariance and competitive exclusion. Mol Ecol 22:2218–2231

Morris GP, Hu Z, Grabowski PP, Borevitz JO, de Graaff M-A, Miller RM, Jastrow JD (2016) Genotypic diversity effects on biomass production in native perennial bioenergy cropping systems. Gcb Bioenergy 8:1000–1014

Mosca E, Eckert AJ, Di Pierro EA, Rocchini D, La Porta N, Belletti P, Neale DB (2012) The geographical and environmental determinants of genetic diversity for four alpine conifers of the European Alps. Mol Ecol 21:5530–5545

Novák FA (1960) Fylogenese serpentinových typů. [Phylogeny of serpentine types.]. Preslia 32:1–8

Nybom H, Bartish IV (2000) Effects of life history traits and sampling strategies on genetic diversity estimates obtained with RAPD markers in plants. Perspect Plant Ecol Evol Syst 3:93–114

Paradis E (2010) pegas: an R package for population genetics with an integrated–modular approach. Bioinformatics 26:419–420

Paradis E, Claude J, Strimmer K (2004) APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20:289–290

Pešout P (2001) Final report: the rescue of Dolnokralovice serpentines and part of Vltava slopes in cooperation with owners. ČSOP, Vlašim

Piry S, Luikart G, Cornuet JM (1999) BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. J Hered 90:502–503

Quintela-Sabarís C, Vendramin GG, Castro-Fernández D, Fraga MI (2010) Chloroplast microsatellites reveal that metallicolous populations of the Mediterranean shrub *Cistus ladanifer* L have multiple origins. Plant Soil 334:161–174

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

Reed DH, Frankham R (2001) How closely correlated are molecular and quantitative measures of genetic variation? A meta-analysis. Evolution 55:1095–1103

Rodríguez-Quilón I, Santos-del-Blanco L, Serra-Varela MJ, Koskela J, González-Martínez SC, Alía R (2016) Capturing neutral and adaptive genetic diversity for conservation in a highly structured tree species. Ecol Appl 26:2254–2266

Russello MA, Waterhouse MD, Etter PD, Johnson EA (2015) From promise to practice: pairing non-invasive sampling with genomics in conservation. PeerJ 3:e1106

Sádlo J, Petřík P, Boublík K, Rychtařík P, Šímová I (2011) Diverzita rostlinstva Hradčanských stěn (Dokesko) a její příčiny [Habitats, vegetation and flora of the Hradčanské stěny rocks (Doksy region, northern Bohemia): causes of diversity]. Zpr Čes Bot Společ 46:17–38

Shafer AB, Wolf JB, Alves PC, Bergström L, Bruford MW, Brännström I, Colling G, Dalén L, De Meester L, Ekblom R (2015) Genomics and the challenging translation into conservation practice. Trends Ecol Evol 30:78–87

Słomka A, Sutkowska A, Szczepaniak M, Malec P, Mitka J, Kuta E (2011) Increased genetic diversity of Viola tricolor L. (Violaceae) in metal-polluted environments. Chemosphere 83:435–442

Smith JM, Smith NH, O'Rourke M, Spratt BG (1993) How clonal are bacteria? Proc Natl Acad Sci 90:4384–4388

Turchetto C, Segatto ALA, Mäder G, Rodrigues DM, Bonatto SL, Freitas LB (2016) High levels of genetic diversity and population structure in an endemic and rare species: implications for conservation. AoB Plants. https://doi.org/10.1093/aobpla/plw002

Wege JA, Thiele KR, Shepherd KA, Butcher R, Macfarlane TD, Coates DJ (2015) Strategic taxonomy in a biodiverse landscape: a novel approach to maximizing conservation outcomes for rare and poorly known flora. Biodivers Conserv 24:17–32

Whitlock R, Hipperson H, Thompson DBA, Butlin RK, Burke T (2016) Consequences of in situ strategies for the conservation of plant genetic diversity. Biol Conserv 203:134–142

Whittaker RH (1954) The ecology of serpentine soils. Ecology 35:258–288

Wolf A (2001) Conservation of endemic plants in serpentine landscapes. Biol Conserv 100:35–44

Wolf AT, Harrison SP (2001) Effects of habitat size and patch isolation on reproductive success of the serpentine morning glory. Conserv Biol 15:111–121

Young A, Boyle T, Brown T (1996) The population genetic consequences of habitat fragmentation for plants. Trends Ecol Evol 11:413–418

Zawko G, Krauss SL, Dixon KW, Sivasithamparam K (2001) Conservation genetics of the rare and endangered *Leucopogon obtectus* (Ericaceae). Mol Ecol 10:2389–2396

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Study 8: Microsatellite markers for study of the invasive species *Rumex alpinus* (Polygonaceae)

Applications in Plant Sciences

PRIMER NOTE

Microsatellite markers for study of the invasive species *Rumex alpinus* (Polygonaceae)

Mária Šurinová[1,2,3], Andrea Jarošová[1], and Zuzana Münzbergová[1,2]

Manuscript received 27 September 2017; revision accepted 28 November 2017.

[1] Institute of Botany of the Czech Academy of Sciences, Zámek 1, 252 43, Průhonice, Czech Republic

[2] Department of Botany, Faculty of Science, Charles University in Prague, Benátská 2, 128 01, Prague, Czech Republic

[3] Author for correspondence: maria.surinova@ibot.cas.cz

Citation: Šurinová, M., A. Jarošová, and Z. Münzbergová. 2018. Microsatellite markers for study of the invasive species *Rumex alpinus* (Polygonaceae). *Applications in Plant Sciences* 6(1): e1017.

doi: 10.1002/aps3.1017

PREMISE OF THE STUDY: Polymorphic microsatellite markers were developed to study genetic diversity, population structure, and dispersal strategies of the highly invasive species *Rumex alpinus* (Polygonaceae).

METHODS AND RESULTS: Fifteen polymorphic microsatellite loci were developed using a 454 sequencing approach and used to genotype 72 individuals from six populations in Austria and France. All markers were polymorphic in at least one investigated population, and the number of alleles ranged from one to four alleles per locus. Observed and expected heterozygosity ranged from 0.06 to 1.0 and from 0.18 to 0.72, respectively.

CONCLUSIONS: These 15 markers will be useful for characterizing dispersal strategies and gene flow assessment between *R. alpinus* populations and other, often weedy *Rumex* species.

KEY WORDS invasive species; microsatellites; Polygonaceae; *Rumex alpinus*.

Rumex alpinus L. (Polygonaceae) is a perennial herb distributed mainly in mountain, subalpine, and alpine zones in Europe. Although the species is considered native in most of Europe, in some areas of Europe it is introduced and considered invasive. These areas include the Orlické Mountains and Krkonoše Mountains in the Czech Republic, the Vosges Mountains in central-western France, and Thuringia and potentially the Black Forest in Germany (Šťastná et al., 2010). It has also been introduced to Great Britain, the United States, and Canada (Šťastná et al., 2010). Populations included in this study were collected in the native range of the species in the French and Austrian Alps. In each population, we collected as many individuals as possible, ensuring that our samples were always at least 2 m apart due to expected clonality of the species. Populations occur in nutrient-rich habitats such as abandoned mountain meadows and along riverbanks (Šťastná et al., 2010). Currently, very little is known about the genetic background of this species. To our knowledge, only one paper has been published to date, using 10 microsatellite markers for phylogenetic reconstruction of five *Rumex* species including *R. alpinus* (Raycheva et al., 2013). These loci were, however, developed for interspecific comparisons and are not suitable for intrapopulation characterization. Despite the invasiveness of this species, no genetic study has been published examining population structure, gene flow between populations, and their relatedness. Although the species is known to be highly clonal (Klimeš, 1992), it is also known to spread both clonally and by seeds (Červenková and Münzbergová, 2009). However, the importance of these two modes of reproduction and their importance for the invasive potential of the species are unknown. We developed and characterized 15 polymorphic markers, which allow for the genetic characterization and further study of this poorly investigated species. Population characterization and estimation of gene flow between populations and an understanding the frequency of these two modes of reproduction will help in designing effective interventions against invasion of this species.

METHODS AND RESULTS

Microsatellite development

Total genomic DNA of 24 individuals (four individuals per population originating from six populations [Appendix 1]) was extracted from dehydrated leaves using a modified cetyltrimethylammonium bromide (CTAB) method (Lodhi et al., 1994). Concentrations were preserved as described in the protocol, but the total volume of the solutions was reduced 10-fold. The sequencing facility GenoScreen (Lille, France) was used to prepare libraries and design primers. Extracted DNA was pooled for microsatellite library preparation. The fragmented DNA was hybridized with eight probes (TG, TC, AAC, AAG, AGG, ACG, ACAT, and ACTC) to enrich the DNA library. The sequencing was performed by a GS FLX sequencer (454 Life Sciences, a Roche Company, Branford, Connecticut, USA). A total of 55,720 reads were obtained and archived in the GenBank Sequence Read Archive (SRR 5822129). QDD software (Meglécz et al., 2010) with default settings was used to identify microsatellite

 © 2018 Šurinová et al. *Applications in Plant Sciences* is published by Wiley Periodicals, Inc. on behalf of the Botanical Society of America. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

loci and for design of the microsatellite primers. A total of 3089 reads contained microsatellite motifs, and 319 candidate microsatellite loci were identified (Appendix S1) with an average sequence length of 325 bp. Markers belonged to di-, tri-, tetra-, penta-, and hexanucleotide repeats (38.28%, 56.3%, 3.5%, 1.28%, and 0.64%, respectively). Across all candidate loci, 2608 primer pairs (3–15 primer pairs per locus) were designed using the Primer3 algorithm implemented within QDD (Malausa et al., 2011), with amplicon lengths ranging between 90 and 320 bp. For each microsatellite candidate locus, one primer pair was selected for further analysis. Of these, we tested 201 primer pairs (Appendix S1) recommended by GenoScreen to identify polymorphic markers. Primers were synthesized (Sigma-Aldrich, St. Louis, Missouri, USA) with M13 tails preceding the 5′ end of the forward primer sequences (Schuelke, 2000). Four individuals from each of six populations of *R. alpinus* (Appendix 1) were used to test amplification efficiency and polymorphism. DNA amplification was performed in 5-μL reactions consisting of 2.5 μL of QIAGEN Multiplex PCR Master Mix, three types of oligonucleotides (0.125 μL of M13-tagged forward primer, 0.125 μL of species-specific reverse primer, and 0.125 μL of fluorolabeled [5′-FAM] M13 primer [10 μM each in initial volume]), 10 ng of DNA dissolved in 0.5 μL TE buffer, and 1.625 μL of H_2O.

The following PCR protocol (Schuelke, 2000) was performed using an Eppendorf Mastercycler Pro S Thermal Cycler (Eppendorf, Hamburg, Germany): an initial denaturation step at 95°C for 15 min; followed by 25 cycles of denaturation (95°C for 20 s), annealing (59°C for 30 s), and extension (72°C for 20 s); followed by 10 cycles of denaturation (95°C for 30 s), annealing (53°C for 45 s), and extension (72°C for 45 s); and a final extension at 72°C for 15 min. During the first 25 cycles specific PCR products are produced, and in the following 10 cycles the fluorescent M13 tag is ligated to the M13 forward primer. A total of 189 primer pairs (94%) were successfully amplified. The majority (74.1%) of loci were monomorphic, and only 49 primer pairs (25.9%) were selected for detailed variability screening on five individuals of *R. alpinus* and five individuals from each of five closely related species (*R. aquaticus* L., *R. crispus* L., *R. obtusifolius* L., *R. sanguineus* L., *R. stenophyllus* Ledeb.). A total of 30 individuals were screened. Based on the variability screening, we chose 15 polymorphic primer pairs for *R. alpinus* species. Most of these were also polymorphic for the closely related species used in the study. GenBank accession numbers for these loci are provided in Table 1.

Genotyping

Total DNA was extracted from 72 *R. alpinus* individuals from six populations and from 46 individuals from five closely related taxa (Appendix 1) for initial primer screening. Primer sequences, repeat motifs, and PCR product size ranges are given in Table 1. PCR reactions were performed as described above. Each PCR product (1 μL) was mixed with 11 μL of a 120 : 1 solution of formamide : size standard (GeneScan 500 LIZ; Thermo Fisher Scientific, Waltham, Massachusetts, USA). Fragment lengths were determined by capillary gel electrophoresis with an ABI 3130 Genetic Analyzer using

TABLE 1. Characteristics of 15 polymorphic loci designed for *Rumex alpinus* genotyping.

Locus[a,b]	Primer sequences (5′–3′)	Repeat motif	PCR product size range (bp)	GenBank accession no.
Ram32	F: AGTAATATTCTTAGACCATCACCACA R: TGTCGACATGAAGCTGTGAA	$(ACCAC)_6$	111–156	MF460423
Ram49	F: GCAGAAGCGGAGGAAGATAC R: CCGCATCGCTACGAGACT	$(GAG)_7$	106–112	MF460424
Ram92	F: TGTGGAGTAAATTGGGGAATG R: GCCCAGCAGTAGCAGTGTC	$(AG)_{14}$	117–127	MF460425
Ram94	F: GGGAGAGGGTTTGGAGATTC R: CACAGGCTCACACGTTGTCT	$(GA)_{17}$	117–137	MF460426
Ram103	F: GTAAACCGCGGCGTAATGTA R: GAGATTCATAGGTGGCGAGG	$(GA)_{11}$	103–107	MF460427
Ram116	F: CAAACAAAAGCCAAACCCAT R: TCTTCATTTCTCCCTCGATCTC	$(GA)_7$	198–202	MF460428
Ram130	F: ATTGAAACCCAGGACCATCA R: TGAAGTGCAACCATAAAGCG	$(AG)_9$	184–190	MF460429
Ram146	F: ACCTTAAACCCTGCCCAAAT R: TCACAACGGTCCAAGTCAAA	$(CAA)_5$	302–317	MF460430
Ram165	F: TGCTACACGAGCTGATGCTT R: TTTCTCTGCTCGGATTGGAT	$(TTC)_6$	149–152	MF460431
Ram168	F: TCGGATTCTTCGAGAGGAGA R: ATTTCCGATTCCCAATTTCC	$(AG)_6$	107–109	MF460432
Ram173	F: AAGGAGCGTGTTTGCAGATAA R: TCAAGTGGGTTTTGTAGCTGTG	$(AG)_5$	113–115	MF460433
Ram193	F: GAGTTCCCACCACAAGAAGC R: CTAGTCCTGGCCTCTACTGC	$(TC)_{10}$	173–189	MF460434
Ram198	F: AATGGGATGGATTGGAAGGC R: ACTCTCACAGTCACAGGCTC	$(GA)_{17}$	147–169	MF460435
Ram199	F: CCCACGAGTTACCCAATCAAC R: TTCTTGACCCTGAATTGTTCG	$(CA)_{17}$	165–177	MF460436
Ram200	F: AAACTGGCCACATTGAGGT R: GTGTCACACTGAAGAACCTG	$(AG)_{14}$	238–248	MF460437

[a]Optimal annealing temperature was 59°C for all loci.
[b]M13 tag (Schuelke, 2000) addition to 5′ terminus.

© 2018 Šurinová et al.

TABLE 2. Results of initial primer screening of 15 microsatellite loci developed for *Rumex alpinus*.[a]

	France (Crozet) (n = 16)			Austria (Hinterstoden Hoss) (n = 5)			Austria (Spital am Pyhrn) (n = 5)			Katschberg (n = 15)			Heiligenblut (n = 15)			Ferleiten (n = 16)		
Locus	A	H_o	H_e	A	H_o	H_e	A	H_o	H_e	A	H_o	H_e	A	H_o	H_e	A	H_o	H_e
Ram32	3	0.37	0.32	2	0.2	0.42	2	0.2	0.42	2	0.26	0.49	2	0.26	0.49	3	0.5	0.44
Ram49	2	1	0.5*	2	1	0.5	2	1	0.5	2	1	0.5*	2	1	0.5*	2	1	0.5*
Ram92	3	0.31	0.43	3	0.6	0.54	3	0.2	0.46	2	0.0	0.44*	2	0.2	0.18	3	0.81	0.65
Ram94	4	0.31	0.52*	1	0.0	0.00	2	0.0	0.48	3	0.26	0.56	2	0.2	0.42	2	0.37	0.43
Ram103	2	0.25	0.30	2	0.4	0.48	2	0.4	0.32	2	0.0	0.49*	2	0.53	0.48	4	0.62	0.62
Ram116	2	0.25	0.37	2	0.6	0.5	2	0.6	0.5	2	0.0	0.49*	2	0.13	0.39	3	0.31	0.35
Ram130	2	0.31	0.49	1	0.0	0.00	2	0.0	0.32	1	0.0	0.0	1	0.0	0.0	1	0.0	0.0
Ram146	3	0.25	0.54	2	0.6	0.42	2	0.4	0.48	1	0.0	0.0	1	0.0	0.0	1	0.0	0.0
Ram165	2	0.06	0.48*	1	0.0	0.00	1	0.0	0.00	1	0.0	0.0	1	0.0	0.0	1	0.0	0.0
Ram168	1	0.00	0.00	2	0.6	0.42	2	0.2	0.18	2	0.66	0.44	1	0.0	0.0	1	0.0	0.0
Ram173	2	0.12	0.49*	1	0.0	0.00	1	0.0	0.00	1	0.0	0.0	1	0.0	0.0	1	0.0	0.0
Ram193	1	0.00	0.00	4	0.0	0.72*	3	0.6	0.46	3	0.2	0.54*	2	0.26	0.49	3	0.75	0.63
Ram198	4	0.31	0.52	1	0.0	0.00	2	0.0	0.48	3	0.26	0.56	2	0.2	0.42	2	0.38	0.43
Ram199	3	0.31	0.35	3	0.6	0.58	2	0.2	0.18	3	0.26	0.24	2	0.26	0.49	3	0.87	0.64
Ram200	3	0.31	0.43	3	0.6	0.54	3	0.2	0.46	2	0.00	0.44*	2	0.2	0.18	3	0.81	0.65

[a]Collection locality and voucher information are provided in Appendix 1.
*Significant deviation from Hardy–Weinberg equilibrium.

TABLE 3. Allele size ranges obtained during cross-amplification trials of microsatellite loci isolated from *Rumex alpinus* and tested in five additional taxa.

Locus	*Rumex crispus*	*Rumex obtusifolius*	*Rumex aquaticus*	*Rumex sanguineus*	*Rumex stenophyllus*
Ram32	96–116	91–146	91–136	91	96–116
Ram49	103–112	106–112	106–109	109	103–124
Ram92	105–111	103–135	87–105	63-105	101–109
Ram94	113–131	123–131	91–127	121	113–127
Ram103	97	107	97–105	267	97
Ram116	198	198–200	184–198	—	198
Ram130	178	178–190	178	178	178
Ram146	299–320	299–323	290–317	317	317–323
Ram165	137–149	146–152	137–155	146	137–149
Ram168	105–107	105–107	105–109	105	105
Ram173	111–115	113–115	111–113	113	111–113
Ram193	159–171	171–193	169–173	173	159–171
Ram198	149–157	153–161	159	149	149–159
Ram199	139–147	159–191	147–175	151–177	147–175
Ram200	226–232	226–256	226	226	226

Note: — = unsuccessful amplification.

GeneMapper 4.0 (Thermo Fisher Scientific). Using GenAlEx 6.5 (Peakall and Smouse, 2012), we calculated observed heterozygosity, expected heterozygosity, and number of alleles. Number of alleles ranged from one to four and observed and expected heterozygosities ranged from 0.06 to 1.0 and from 0.18 to 0.72, respectively. GENEPOP 4.2 (Rousset, 2008) was used to calculate deviations from Hardy–Weinberg equilibrium and linkage disequilibrium. Significant linkage disequilibrium was found within four populations following Bonferroni correction (Crozet population: loci 94/198 and 165/173; Katschberg population: loci 198/94, 193/94, 198/32, and 94/32; Heiligenblut population: loci 199/32, 193/32, and 198/94; Ferleiten population: 198/94 and 200/92). Significant deviation from Hardy–Weinberg equilibrium was found mainly in the Crozet and Katschberg populations (Table 2), which indicates strong clonality. This indication is also supported for the Crozet population by lower observed heterozygosity than expected for each locus of this population (Table 2).

We also tested cross-amplification of these loci in five other *Rumex* taxa: *R. aquaticus*, *R. crispus*, *R. obtusifolius*, *R. sanguineus*, and *R. stenophyllus*. Tests for cross-amplification and polymorphism in the five congeneric taxa resulted in successful genotyping of up to 14 of the 15 loci (Table 3). These results demonstrate that these primer pairs may be of broad utility throughout the genus *Rumex*.

CONCLUSIONS

We developed and successfully amplified 15 polymorphic markers in several *Rumex* taxa, including *R. alpinus*, *R. aquaticus*, *R. crispus*, *R. obtusifolius*, *R. sanguineus*, and *R. stenophyllus*. These polymorphic loci will be valuable for the future study of *R. alpinus* population genetic composition and its influence on the invasiveness of the species.

ACKNOWLEDGMENTS

This work was supported by the Charles University Grant Agency (grant no. 1522214).

© 2018 Šurinová et al.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

LITERATURE CITED

Červenková, Z., and Z. Münzbergová. 2009. Susceptibility of the landscape of the Giant Mountains, Czech Republic, to invasion by *Rumex alpinus*. *In* P. Pyšek and J. Pergl [eds.], Biological invasions: Towards a synthesis. *Neobiota* 8: 75–85.

Klimeš, L. 1992. The clone architecture of *Rumex alpinus* (Polygonaceae). *Oikos* 63: 402–409.

Lodhi, M. A., G. N. Ye, N. F. Weeden, and B. I. Reisch. 1994. A simple and efficient method for DNA extraction from grapevine cultivars and *Vitis* species. *Plant Molecular Biology Reporter* 12: 6–13.

Malausa, T., A. Gilles, E. Meglécz, H. Blanquart, S. Duthoy, C. Costedoat, V. Dubut, et al. 2011. High-throughput microsatellite isolation through 454 GS-FLX Titanium pyrosequencing of enriched DNA libraries. *Molecular Ecology Resources* 11: 638–644.

Meglécz, E., C. Costedoat, V. Dubut, A. Gilles, T. Malausa, N. Pech, and J. F. Martin. 2010. QDD: A user-friendly program to choose microsatellite markers and design primers from large sequencing projects. *Bioinformatics* 26: 403–404.

Peakall, R., and P. E. Smouse. 2012. GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research–an update. *Bioinformatics* 28: 2537–2539.

Raycheva, T., I. Denev, and D. Dimitrova. 2013. Taxonomic relationships of selected Bulgarian species from *Rumex* subg. *Rumex* (Polygonaceae) based on ISSR markers. *Phytologia Balcanica* 19: 29–37.

Rousset, F. 2008. GENEPOP'007: A complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources* 8: 103–106.

Schuelke, M. 2000. An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology* 18: 233–234.

Šťastná, P., L. Klimeš, and J. Klimešová. 2010. Biological flora of Central Europe: *Rumex alpinus* L. *Perspectives in Plant Ecology, Evolution and Systematics* 12: 67–79.

APPENDIX 1. Accession information for *Rumex* species in this study.[a]

Species	Collection locality/ Population code	Country	*n*	Latitude	Longitude
R. alpinus L.	Crozet	France	16	46.2866339	6.0107875
R. alpinus	Hinterstoden Hoss	Austria	5	47.696796	14.153384
R. alpinus	Spital am Pyhrn	Austria	5	47.658855	14.354099
R. alpinus	Katschberg	Austria	15	47.055959	13.616254
R. alpinus	Heiligenblut	Austria	15	47.064204	12.827788
R. alpinus	Ferleiten	Austria	16	47.134005	12.807474
R. aquaticus L.	Markvartice	Czech Republic	10	50.4294492	15.1982517
R. crispus L.	Markvartice	Czech Republic	10	50.4294492	15.1982517
R. obtusifolius L.	Markvartice	Czech Republic	10	50.4294492	15.1982517
R. sanguineus L.	Markvartice	Czech Republic	6	50.4294492	15.1982517
R. stenophyllus Ledeb.	Černolice	Czech Republic	10	49.91623	14.30901

Note: n = number of individuals.

[a]Leaf samples of 5–16 individuals per population were collected in the field and dried (silica gel) before performing DNA extraction. Herbarium vouchers, leaf samples, and DNA extracts are deposited at the Institute of Botany of the Czech Academy of Sciences (PRA), Průhonice, Czech Republic.

© 2018 Šurinová et al.

3. Conclusions

This thesis provides some insights into the population structure of polyploid species and species with fragmented populations.

Population differentiation in the diploid- tetraploid complex of *Arabidopsis arenosa* with limited gene flow showed expected selection on meiotic genes and found differences in meiotic stabilization between both cytotypes. The diploid- hexaploid complex of *Aster amellus* was studied with general phylogenetic markers and microsatellite regions. We identified strongly restricted gene flow between cytotypes even at the contact zone, and we proposed that these cytotypes may be considered as two different lineages. In both cases, polyploidization plays a major role in population differentiation and in the case of *Aster amellus*, this differentiation led to the establishment of new (cryptic) species.

Temperature and precipitation may be important determinants for population differentiation of hexaploid species *Festuca rubra*, but ploidy variation does not play any role in the system. We found no significant relationship between genetic and geographic distance matrices for the total data set. However, genetic distances were significantly positively related to the environmental distances. The environmental distance combining temperature and precipitation have the strongest relationship, and precipitation distance has the weakest relationship. Most of the genetic variation was distributed within populations, but an increase in clonality was found for the alpine population. Also, precipitation factor alone and in interaction with temperature strongly influenced population genotypic diversity suggesting increased clonality towards the coldest and especially the coldest wettest climates. Thus, in this study we showed, that genetic differentiation of the populations is significantly linked to the environment of origin.

For characterization of *Gentianella praecox subsp. bohemica* populations were necessary to develop species-specific microsatellite markers. For that, 454 sequencing was used, and we designed multiplex PCR reactions with 12 polymorphic regions, and 1400 individuals from 30 populations were genotyped. We surprisingly found that large reductions in population size do not result in immediate loss of genetic diversity and reduction of fitness within the populations. This is likely due to the seed bank of the species serving as a reservoir of the genetic diversity of the populations.

The second endangered species which we characterized was *Minuartia smejkalii*. SNPs from the remaining seven populations obtained by nextRAD sequencing showed high genetic diversity within *M. smejkalii* populations, low to moderate among-populations differentiation, and moderate regional differentiation. Also, 2–3% of the markers show differentiation patterns consistent with divergent selection, suggesting that some local adaptation might have occurred.

For species *Rumex alpinus*, we also developed a set of species-specific microsatellite regions with 454 sequencing. We were able to find 15 polymorphic regions, but initial screening of 6 populations showed strong clonal propagation and small differences between populations. Clonal growth is thus a preferred reproduction strategy and contributes greatly to the dispersal ability of this species.

For both endangered species, we hypothesized that genetic variability is preserved thanks to reproduction strategies. In the case of *Gentianella praecox subsp. bohemica* thanks to seed bank reservoir, in case *Minuartia smejkalii as* a consequence of an outcrossing mating system and long life cycle. In contrast, the preferred clonal strategy led to the low population genetic variability of invasive species *Rumex alpinus*.