



*Łukasz Walas**, *Grzegorz Iszkuło*, *Zoltan Barina*, *Monika Dering*

Development of microsatellite markers for horse-chestnut (*Aesculus hippocastanum*), their polymorphism in natural Greek populations, and cross-amplification in related species

Received: 15 March 2021; Accepted: 22 April 2021

Abstract: New nuclear microsatellite markers (SSRs) were developed for *Aesculus hippocastanum*, a relict tree species from the Balkan Peninsula. The development of microsatellites was done using the Illumina MiSeq PE300 platform. Out of a set of 500 SSRs designed, a subset of 13 loci was tested using 290 individuals from seven natural populations. Twelve species-specific loci were polymorphic. The number of alleles per locus ranged from 2 to 17 and expected heterozygosity from 0.089 to 0.800 with a mean value of 0.484. The population of Kalampaka had the lowest value of allelic richness (2.63) and gene diversity in comparison to the remaining populations. STRUCTURE analysis confirmed isolation of population Mariolata from the southern edge of the species range and genetic similarity among populations from the Pindos Mts. Additionally, the utility of new SSRs in 29 individuals from nine other *Aesculus* taxa was tested. Eleven markers gave polymorphic products for all tested species. For 24 individuals, a high-quality product was obtained for each marker. Results confirmed the utility of specific markers for future population genetics studies.

Keywords: Tertiary relict, endemic species, cross-amplification, polymorphism

Addresses: Ł. Walas, G. Iszkuło, M. Dering, Institute of Dendrology, Polish Academy of Sciences, Parkowa 5, 62-035 Kórnik, Poland, e-mail: lukaswalas@man.poznan;

ŁW <https://orcid.org/0000-0002-4060-9801>

G. Iszkuło, Institute of Biological Sciences, University of Zielona Góra, Prof. Z. Szafrana 1, 65-516 Zielona Góra, Poland, e-mail: iszkulo@man.poznan.pl; <https://orcid.org/0000-0003-2067-729X>

Z. Barina, Department of Botany, Hungarian Natural History Museum, H-1431 Budapest, Pf. 137, e-mail: barina.zoltan@nhmus.hu; <https://orcid.org/0000-0003-3117-7186>

M. Dering, Faculty of Forestry, Poznań University of Life Sciences, Wojska Polskiego 85, 60-637 Poznań, Poland, e-mail: mdering@man.poznan.pl; <https://orcid.org/0000-0001-7017-5541>

*Corresponding author

Introduction

Global climate changes are projected to affect the whole biosphere, by disturbing the functioning of ecosystems and shifting species distributions (Peñuelas et al., 2013; Pecl et al., 2017). Many endemic and relict taxa may not be able to withstand these changes, and their extinction could cause a

significant decline in biodiversity (Casazza et al., 2014). A study shows that stable climatic conditions support high endemism (Harrison & Noss, 2017), a factor that may no longer be valid in some regions soon. The Mediterranean hotspot of diversity is one of the most important areas in terms of plant diversity in Europe – about 25,000 species occur in this region and half of them are endemics (Myers et al.,

2000). Predictions for the future climate in the Mediterranean area show an increase in temperatures, reduction in rainfalls, and higher variance in seasonal patterns of precipitation (Giorgi & Lionello, 2008), which can pose a serious threat to many species, particularly endemics (Orsenigo et al., 2018). Considering the rapid pace of the current climate changes, more studies are needed to support setting up strategies for the efficient conservation of endemic trees to prevent their loss and support their survival.

Horse-chestnut (*Aesculus hippocastanum* L.) is a Tertiary relict and endemic species that occurs in the mountainous regions of the Balkan Peninsula. This is the only representative of the genus *Aesculus* in Europe. The species is well-known because of its ornamental value, which was the reason for its great popularity in urban flora across Europe and worldwide (Lack, 2000). But despite that, natural populations of this species are currently endangered because of climate changes, human activities, as well as the spread of diseases and pests (Thalmann et al., 2003; Steele et al., 2010; Jagiełło et al., 2017; Walas et al., 2018). The International Union for Conservation of Nature (IUCN) recommends urgent genetic research on natural stands of horse-chestnut (Allen & Khela, 2017). Analysis of genetic diversity and structure can help to understand the interplay between demographic processes and selection which occur in the natural populations of this plant. Knowledge about the level and the spatial patterns of genetic diversity is crucial for planning an efficient in situ conservation strategy. Genetic markers, which allow describing population genetic structure with high accuracy, are a vital help in prioritizing the conservation actions. The level of genetic diversity informs on how the species reacted to environmental factors in the past and what we can expect in the future in terms of its adaptive capacity. For that reason, data on genetic diversity is a prerequisite for designing a sound conservation strategy for horse-chestnut. However, genetic information must be generated using reliable methods that deliver good quality data to properly conclude on species genetic resources conserved in natural stands.

Despite fast technological progress, microsatellite markers (SSRs) remain one of the most important molecular ecology tools because of their easy allele detection, high polymorphism, and relatively low costs (Vieira et al., 2016; Mora et al., 2017). For small populations with low migration rates, as in the case of most endemic species, even a small number of microsatellites can provide significant information on microevolutionary processes, their rate, and directions (Selkoe & Toonen, 2006). Due to the lack of specific microsatellites for *A. hippocastanum*, the markers developed for the closely related Japanese horse-chestnut, *A. turbinata* Blume were used for the

assessment of the population genetic structure (Minami et al., 1998; Walas et al., 2019). Many SSRs show a high rate of transferability between close taxa; this has also been proven for several tree species and genera (Vignes et al., 2006; Ravishankar et al., 2011; Boratyński et al., 2014). However, cross-amplification can give low-quality products, which may cause scoring errors. Additionally, reaction failures that may indicate the existence of interspecies sequence differences in the flanking DNA regions, which are targets for primers, may lead to a substantial null allele frequency, and in consequence, affect estimators of intra- and interpopulation diversity, leading to biased inferences. A locus that is polymorphic in one species may be monomorphic in another, or the products may be non-specific (Sugai et al., 2016; Godoy et al., 2019). Therefore, the development of new, polymorphic markers specific to the *A. hippocastanum* genome is necessary for accurate characterisation of the population structure and patterns of genetic diversity in remnant natural stands of the species.

In our work, we aimed to characterise a set of new polymorphic markers designed for *A. hippocastanum* that could be used in diversity studies with conservation perspectives. Three multiplex reactions for 12 novel SSRs were optimised and tested in a study of the population genetic structure of natural populations of the species from Greece. Additionally, we used STRUCTURE analysis to check, whether the size of the loci set might affect the detection of the population structure. For this purpose, we used species-specific and cross-amplified loci used in the previous study (Walas et al., 2019) to enlarge the marker set and evaluate its resolution power. Finally, we aimed to deliver information about the potential utility of these newly designed SSRs in other *Aesculus* taxa.

Materials and methods

The development of novel microsatellites was performed at AllGenetics (www.allgenetics.eu). One sample was used to generate a library with the Nextera XT DNA Library Preparation Kit (Illumina). The library was enriched with fragments with microsatellite motifs by hybridisation to four groups of oligopeptides (AG, AC, ATCT, and ACG) and was sequenced in the Illumina MiSeq PE300 platform (Macrogen Inc.). The library produced 6,990,226 sequences. Reads were processed in Geneious 10.2.3 (Biomatters Ltd). Primer design was carried out in Primer 3 software (Koressaar & Remm, 2007; Untergasser et al., 2012), implemented in Geneious 10.2.3. Finally, 500 primer pairs were developed during the procedure (Table S1). These primer pairs are located at

the flanking regions of the microsatellite motifs. Forty primer pairs were randomly chosen for the next step of the procedure. These pairs were multiplexed in the sets of three to five markers, based on their features, and tested in eight individuals. In the last step, 10 primers were tested in three individuals, to optimize the final PCR reaction.

A total number of 13 primer pairs – 10 tested in AllGenetics and three additional (AH_037, AH_051, AH_054, AH_101, AH_129, AH_222, AH_257, AH_269, AH_359, AH_375, AH_419, AH_447, and AH_485) were organised into three sets according to their properties and expected amplicon sizes (Tables 1 and 2). These markers were tested on 290 individuals of *A. hippocastanum* originating from seven natural populations from Greece (Table 3) and used in the previous study (Walas et al., 2019). DNA was extracted from leaves according to the protocol described by Dumolin et al. (1995). PCR reactions were conducted in a volume of 10 µL, containing 1 × reaction buffer, 0.1 µg of BSA (Bovine Serum Albumin), 1.5 mM MgCl₂, 2 µM of dNTP mix, 0.5 U of SilverTaq polymerase (Syngen, Poland), 0.05 µM of each starter and 100 ng of DNA. Reactions were conducted using the following protocol: initial denaturation at 95 °C for 12 min, followed by 30 cycles of denaturation at 95 °C for 30 s, annealing at 57 °C for 90 s, elongation at 72 °C for 30 s; 8 cycles of 95 °C for 30 s, 53 °C for 90 s, 72 °C for 30 s and final elongation at 68 °C for 15 min. Products of amplification were analysed using 3130 Genetic Analyser (Applied Biosystems, Foster City, California, USA) with internal size standard GeneScan LIZ-500. Genotypes were scored using GENEMAPPER v. 4.0 (Applied Biosystems). Amplification and genotyping for eight loci developed for *A. turbinata* (AT3D6, AT6D8, AT7D1, AT5D2, AT6D11, AT6D2, AT7D8, and AT6D12) were conducted during the previous study (Walas et al., 2019).

New markers were also tested on 29 individuals representing nine taxa from the genus *Aesculus*

collected from the Adam Mickiewicz University Botanical Garden in Poznań and from Kórnik Arboretum of the Institute of Dendrology PAS. The collected individuals were as follows: *A. ×carnea* Hayne (3 individuals), *A. chinensis* Bunge (2 individuals), *A. flava* Sol. (4 individuals), *A. glabra* Willd. (7 individuals), *A. ×hybrida* DC. (1 individual), *A. ×neglecta* Lindl. (4 individuals), *A. parviflora* Walter (3 individuals), *A. pavia* L. (3 individuals), and *A. turbinata* Blume (2 individuals).

Basic diversity estimates such as an average number of alleles (N_a) and an effective number of alleles (N_e) were estimated using GENEALX 6.4 (Peakall & Smouse, 2006). FSTAT v. 2.9.3 (Goudet, 2001) was used to calculate allelic richness (A_r), while INEST v. 2.0 (Chybicki, 2016) was applied for calculating genotyping error rate (b), observed (H_o) and expected (H_e) heterozygosity, as well as inbreeding coefficient (F_{is}). Wright's fixation index (F_{ST}) was estimated in FREENA, with and without Excluding Null Alleles (ENA) correction (Chapuis & Estoup, 2007). The latter software was also used to estimate the frequency of the null alleles ($Null$). A test of

Table 2. Arrangement of 13 microsatellite markers into multiplex reactions

Loci	Multiplex	Dye	N
AH_051	I	PET	14
AH_054	I	VIC	7
AH_269	I	FAM	2
AH_359	I	FAM	7
AH_447	I	NED	17
AH_101	II	NED	2
AH_129	II	PET	5
AH_037	II	FAM	6
AH_419	II	VIC	6
AH_222	III	FAM	3
AH_257	III	NED	7
AH_375	III	PET	6
AH_485	III	VIC	7

N – number of alleles.

Table 1. Species-specific primers for 13 microsatellites tested in the *Aesculus hippocastanum*

Loci	Forward Primer	Reverse Primer	Size range (bp)	Motif
AH_037	ATTGAACGCAGCTCTCACCT	GTGAACTAAGACCGTGCCGT	68–98	(CTCTC) ₅
AH_051	TCGGGATGTCGTTATTAGGC	CCAGAAATCCGCACTAAGC	280–312	(AT) ₇
AH_054	AGCCTCCTCTCGACGTACAA	CGCCGAAGCTGTTGATTATT	138–162	(TCG) ₇
AH_101	CGCACTCGCAAGTTTGAGTA	TGGTCATCACAGTGGCTGAT	202–205	(TAG) ₅
AH_129	CCTGGAATCTCAACCAGGA	CTTTGGAAGGAGGCATCTCA	104–114	(AT) ₅
AH_222	AGAAGTCTCCAGACCACG	CCTGCAAACAGAATCGGAAT	150–162	(TTC) ₅
AH_257	AATGGTTACTGACGTGGCCT	CGCGTGTCTTAAACGAGTCA	80–98	(CT) ₅
AH_269	TTGTCACGAACATGTCCACC	ATGGTTCACCTTGAACGGG	216–232	(TG) ₅
AH_359	CTCCTCGACCAACAACACTG	AGGACATTCGAGGATGACG	92–128	(CTCCCT) ₅
AH_375	TTACAATCCTCCAACCCTGC	ATATCACCCAGCACCTCCCT	233–257	(CAA) ₅
AH_419	GTTCTCCCTTGGCCTTGACT	GGTAAAGCGAGGTACGCAA	111–129	(CAG) ₈
AH_447	CGGGTAGTAGCACAGCAACA	TTGGAACCAATTCAGAACCG	105–139	(TA) ₇
AH_485	CATGATCGGTTGAACAGTGC	TGTCGAGCAAACACATCAGG	82–106	(GTG) ₆

Table 3. Location of the populations of *Aesculus hippocastanum* used in this study

Population	Voucher	Latitude	Longitude	Altitude	Region	N
Ondria	KOR 51217	40°20'N	21°05'E	1463	Pindos Mts	50
	KOR 51218					
Kalampaka	No voucher	39°48'N	21°16'E	1371	Pindos Mts	23
Dasos Nanitsa	KOR 51216	39°42'N	21°21'E	1029	Pindos Mts	93
Vaeni	No voucher	39°12'N	21°42'E	1089	Pindos Mts	32
Mariolata	KOR 51230	38°37'N	22°26'E	1239	Parnassus Massif	42
	KOR 51219					
Karitsa	KOR 51280	39°48'N	22°45'E	705	Ossa Massif	24
Perivoli	KOR 51226	39°58'N	21°11'E	915	Pindos Mts	26

N – number of individuals sampled

Hardy-Weinberg Equilibrium (HWE) was performed using the *hw.test* function in the package “pegas” and visualised with the function *levelplot* from the package “lattice” in the R environment (Sarkar, 2008; Paradis, 2010; R Core Team, 2013). Genotyping linkage disequilibrium (LD) for each pair of loci was calculated with GENEPOP on the Web application (Raymond, 1995; Rousset, 2008) using likelihood ratio statistics and default Markov chain parameters. For HWE and LD tests, the p-value was corrected using Bonferroni correction. Analysis of Molecular Variance was conducted in GENEALX (Peakall & Smouse, 2006). This software was also applied for the calculation of the Codom-Genotypic Genetic Distance between all tested species. Obtained pairwise distances were visualized with the function *pcoa* in the package “ape” in the R environment (R Core Team, 2013; Paradis & Schliep, 2019).

Individuals were divided into genetic clusters using a non-spatial Bayesian clustering model implemented in STRUCTURE 2.3.4 (Pritchard et al., 2000). The procedure included 10 independent runs with 10^5 of burn-in and 10^6 MCMC iterations with the maximum number of clusters set to $K=8$, correlated allele frequencies within populations assumed, and mixed ancestry of individuals allowed. Three analyses were performed: 1) for species-specific loci only (a set of 12 loci), 2) for cross-amplified loci (from *A. turbinata*; a set of 8 loci) and 3) for combined sets of loci. To estimate the best-supported number of clusters, Evanno’s delta K method implemented in CLUMPAK (Kopelman et al., 2015) was used.

Reserve Selection analysis from DIVA-GIS software (Hijmans et al., 2011) was used to indicate the populations with the priority of conservation. This method estimates the minimum number of geographical units (such as regions or populations) necessary to conserve all genetic diversity of the targeted species. We used alleles obtained for species-specific loci with “Rarity” option, which takes into account the frequency of the alleles. The results of the analysis were visualized in QGIS 3.10.6. (QGIS Development Team, 2012).

Results

Twelve of the tested loci were polymorphic, and the number of alleles ranged between 2 and 17. Locus AH_269 showed the same profile of microsatellite peaks for all tested individuals and thus it was excluded from further analysis due to lack of polymorphism (Table 4, Table S2). The remaining loci showed good interpretable and reproducible polymorphic patterns without visible errors, such as stutter bands or split peaks (Fig. S1). The efficiency of amplification in *A. hippocastanum* was high as we obtained 0–3.1% of missing data, with the average value reaching only 0.9%. The highest missing values were observed for loci AH_419 (3.1%) and AH_375 (2.8%). Additionally, genotyping error (0.07%) was very low in comparison with the typical range noted for microsatellites (Wang, 2018) and was lower than the error rate in cross-amplified markers, for which mistyping was 0.11% (Table S3). Wright’s fixation index was slightly lower for specific markers than for cross-amplified loci (values of F_{ST} with ENA correction were 0.113 and 0.116, respectively). Expected heterozygosity ranged between 0.089 (locus AH_222) and 0.800 (locus AH_447), with a mean value of 0.484. The average frequency of null alleles was 0.062, with values >0.1 noted in loci AH_037 (0.111), AH_419 (0.113) and AH_257 (0.118). These markers should be used with caution, preferably with the correction methods applied in FREENA or another software that accounts for null alleles (Chapuis & Estoup, 2007). According to the analysis performed for a set of 12 loci, some pairs of loci were in significant linkage disequilibrium (Table S4). Specifically, loci AH_051 and AH_257 were associated with four other loci, AH_485 with three loci, while AH_054 and AH_375 were associated with two loci. However, the number of pairs with significant linkage disequilibrium differed in each population analysed separately. In two populations (Kalampaka and Karitsa) all loci were independent. Loci AH_257 and AH_375 did not conform to HWE in five populations and locus AH_447 did not conform in four populations (Fig. 1).

Table 4. Variability of newly designed SSRs markers specific for *Aesculus hippocastanum*

Loci	Na	Ne	Null	AR	H _O	H _E	F _{IS}	F _{ST}	F _{ST} ENA	b
AH_037	4.14	2.11	0.111	4.204	0.329	0.570	0.424	0.233	0.202	0.01%
AH_051	6.43	2.82	0.034	8.211	0.573	0.733	0.218	0.203	0.200	0.04%
AH_054	4.71	2.39	0.004	5.172	0.627	0.638	0.018	0.141	0.139	0.59%
AH_101	2.00	1.45	0.075	2.140	0.190	0.325	0.415	0.179	0.172	0.01%
AH_129	2.57	1.25	0.034	2.377	0.148	0.154	0.035	0.140	0.137	0.01%
AH_222	2.00	1.10	0.008	1.937	0.086	0.089	0.032	0.012	0.011	0.01%
AH_257	4.71	2.74	0.118	5.333	0.429	0.681	0.370	0.046	0.056	0.01%
AH_359	4.43	2.54	0.053	4.501	0.516	0.648	0.204	0.107	0.101	0.01%
AH_375	3.57	1.42	0.083	4.054	0.202	0.332	0.390	0.042	0.064	0.03%
AH_419	3.29	1.65	0.113	3.782	0.224	0.435	0.484	0.044	0.054	0.02%
AH_447	7.43	3.74	0.068	9.357	0.668	0.800	0.165	0.137	0.135	0.02%
AH_485	3.57	1.64	0.048	4.857	0.363	0.405	0.103	0.106	0.087	0.02%
Average	4.08	2.07	0.062	4.660	0.363	0.484	0.238	0.116	0.113	0.07%

The average number of alleles in a population (N_a) ranged from 2.67 in Kalampaka to 4.92 in Karitsa I, whereas the average effective number of alleles (N_e) varied from 1.74 in Kalampaka to 2.26 in Vaeni (Table 5). In the population of Kalampaka, loci AH_222 and AH_375 were monomorphic. Average allelic richness (AR) was 3.71 and was similar in all populations except for Kalampaka, where it was

much lower (2.63). Genetic diversity (H_e) ranged from 0.345 in Kalampaka to 0.495 in Perivoli, while observed heterozygosity (H_o) varied from 0.326 in Kalampaka to 0.394 in Ondria. The average number of private alleles was 3.14. Totally, 22 private alleles were detected which is much more than in the previous study with SSRs loci designed for *A. turbinata* in which 13 private alleles were observed (Walas et al., 2019). AMOVA showed that 13% of molecular variance occurs among populations, 16% among individuals, and 71% within individuals (Table S5).

The new SSRs were successfully amplified in all tested taxa and gave polymorphic and high-quality products. For one individual of *A. chinensis*, we did not obtain the products for AH_037, AH_101, and AH_419 loci. Primer AH_051 did not give the product in one individual of *A. flava*, one of *A. glabra*, one of *A. ×neglecta*, and one of *A. parviflora*. Additionally, 18 alleles not present in *A. hippocastanum* were observed in other species (Table S2). Eight alleles were detected in *A. chinensis* – of which two were presented only in this species (one for AH_051 and one for AH_054 markers). Seven alleles not detected in *A. hippocastanum* were noted in *A. flava*, eight in *A. glabra* (two alleles for AH_129 were unique for this species), eight in *A. ×hybrida* (one allele of AH_222 was detected only in this taxa), eleven in *A. ×neglecta*, two for *A. parviflora*, one in *A. turbinata* and five in *A. pavia*. All alleles detected in *A. ×carnea* were presented also in *A. hippocastanum*. Interestingly, locus AG_375 was monomorphic for all species except for *A. hippocastanum*. Despite the results, the usefulness of these SSRs for initial species identification should be verified with a greater number of individuals. Codom-Genotypic Genetic Distance between all tested species showed results in accordance with the current taxonomy (Fig. 2, Table S6).

Analysis of genetic structure based on 20 loci (12 specific for *A. hippocastanum* and 8 specific for *A. turbinata*) made with STRUCTURE (Pritchard et al., 2000) defined seven genetic clusters, which clearly

Locus	1	2	3	4	5	6	7
AH_375	0.0000	M	0.0000	0.0000	0.0000	0.7314	0.0002
AH_222	0.4231	M	0.0034	0.9285	0.7459	0.7440	0.7549
AH_257	0.0000	0.0452	0.0000	0.0000	0.0000	0.0000	0.3499
AH_485	0.2991	0.1444	0.0000	0.1348	0.0000	0.1057	0.2589
AH_037	0.8400	0.0003	0.0000	0.0009	0.0276	0.2099	0.7841
AH_419	0.0000	0.6479	0.0000	0.0000	0.1052	0.9480	0.3785
AH_101	0.0007	0.7379	0.0484	0.0054	0.8744	0.8044	0.0011
AH_129	0.1780	0.0565	0.9975	0.7809	0.9378	0.0000	0.0000
AH_359	0.0000	0.5753	0.5715	0.0000	0.0488	0.1207	0.0192
AH_051	0.8532	1.0000	0.0000	0.0186	0.4954	0.8800	0.2292
AH_054	0.9401	0.6587	0.8375	0.9532	0.1465	0.5491	0.5321
AH_447	0.0000	0.1655	0.0000	0.0071	0.0000	0.4906	0.0000

Fig. 1. Results of the test of Hardy-Weinberg equilibrium (HWE) according to loci and populations (p-values were corrected using Bonferroni correction). Significant departures from HWE are indicated by grey color. "M" indicates that locus is monomorphic in a given population
Population numbers: 1 – Ondria, 2 – Kalampaka, 3 – Dasos Nanit-sa, 4 – Vaeni, 5 – Mariolata, 6 – Karitsa, 7 – Perivoli.

Table 5. Parameters of genetic diversity of the studied populations of *Aesculus hippocastanum* for the loci designed in this study

Population	N	Na	Ne	Null	AR	A _p	H _O	H _E	F _{IS}	b
Ondria	50	4.42	2.22	0.068	3.90	3	0.394	0.472	0.030	0.16%
Kalampaka	23	2.67	1.74	0.036	2.63	0	0.326	0.345	0.025	0.09%
Dasos Nanitsa	93	4.92	1.92	0.078	3.88	6	0.346	0.427	0.024	0.23%
Vaeni	32	4.25	2.26	0.074	3.96	2	0.342	0.433	0.124	0.08%
Mariolata	42	4.42	2.08	0.050	3.87	3	0.371	0.411	0.042	0.05%
Karitsa	24	3.92	2.07	0.042	3.83	4	0.392	0.431	0.054	0.42%
Perivoli	26	3.92	2.20	0.086	3.89	4	0.385	0.495	0.065	0.48%
Average		4.07	2.07	0.062	3.71	3.14	0.370	0.430	0.050	0.22%

N – number of individuals, Na – the average number of alleles, Ne – effective number of alleles, Null – frequency of null alleles, AR – allelic richness, AP – number of private alleles, Ho – observed heterozygosity, HE – expected heterozygosity, FIS – inbreeding coefficient, b – genotyping error rate.

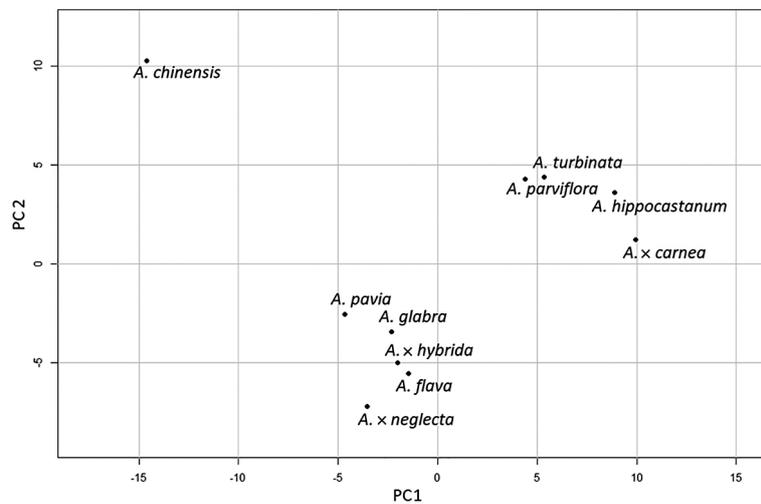
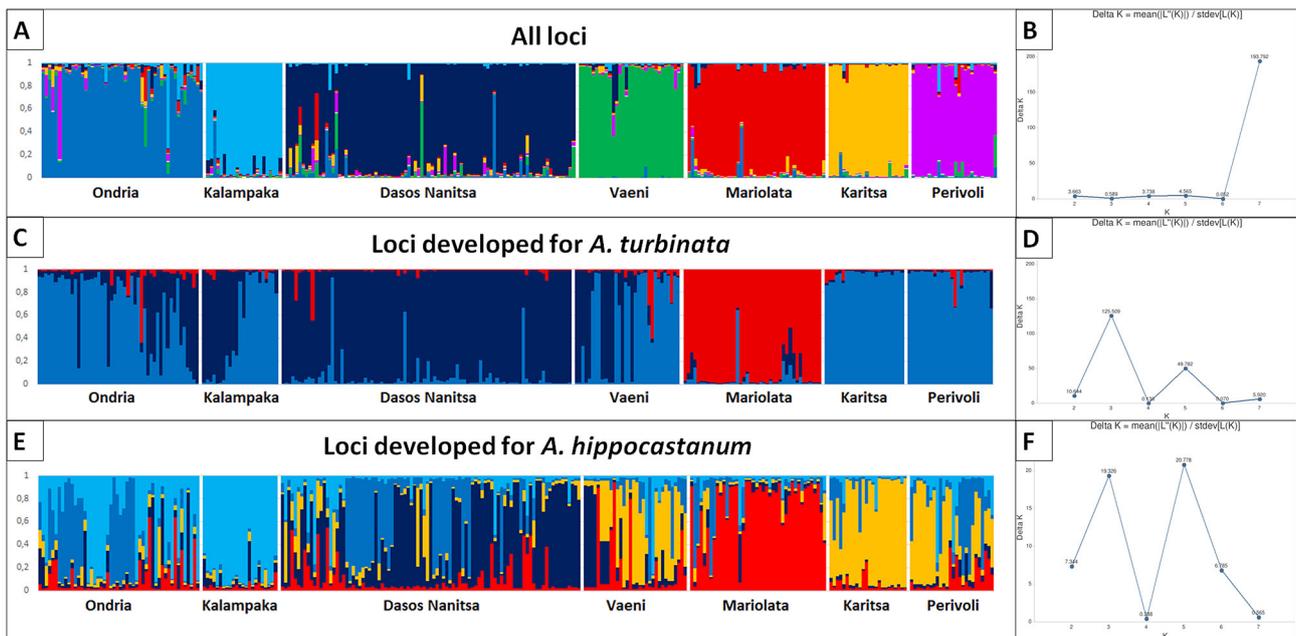
Fig. 2. Results of Principal Coordinate Analysis according to Codom-Genotypic Genetic Distance for tested taxa from genus *Aesculus*

Fig 3. Individuals grouped by population and genetic clustering as a result of the STRUCTURE analysis
 A – clusters for all loci for K=7; B – the best K for all loci; C – clusters for loci developed for *Aesculus turbinata* for K=3; D – the best K for loci developed for *A. turbinata*; E – clusters for loci developed for *A. hippocastanum* for K=5; F – the best K for loci developed for *A. hippocastanum*.

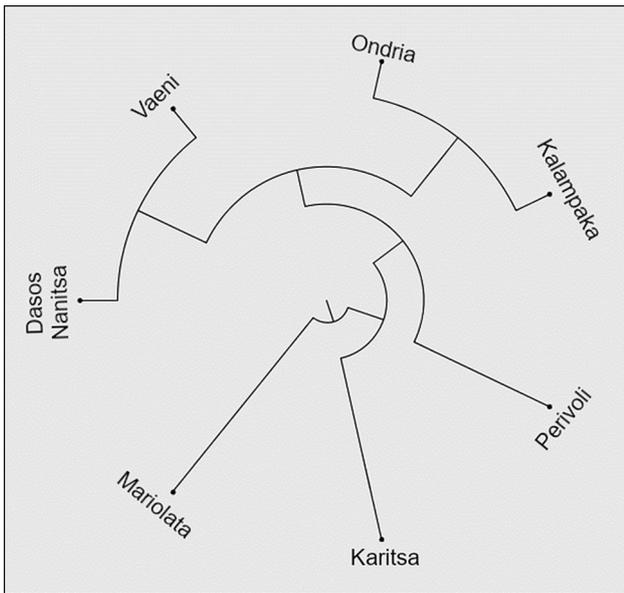


Fig. 4. UPGMA tree based on Nei's genetic distances for natural populations of *Aesculus hippocastanum*

matched to the populations investigated (Fig. 3). Analyses for both sets of loci separately showed $K=3$ and $K=5$ as the most probable for cross-amplified loci and species-specific loci, respectively. However, based on the species-specific loci set, $K=3$ and $K=5$ were almost equally probable. UPGMA clustering (Fig. 4) showed isolation of Mariolata and similarity between populations from the Pindos Mountains, which was also confirmed by values of the pairwise F_{ST} (Table S7). Reserve selection analysis pointed at Dajos Nanitsa, Karitsa and Mariolata as sites that should be protected with high priority (Fig. 5).

Discussion

The new markers, specific for *A. hippocastanum*, showed lower allelic richness than cross-amplified loci from *A. turbinata* (Table 4, Table S3). The expected and observed heterozygosity estimated with specific markers were also lower. This result can be



Fig. 5. Location of the analyzed populations of *Aesculus hippocastanum* (red dots). The size of the dots indicates the conservation priority according to the allelic richness values. Dashed areas show Sites of Community Importance (SCI) for the protection of flora and fauna in Greece (source: geodata.gov.gr)

related to the length of the microsatellite motifs because markers with dinucleotide motifs usually have a higher mutation rate and a higher number of alleles than markers with longer motifs (Zurn et al., 2020). All cross-amplified loci have a dinucleotide motif, whereas as many as eight new markers have a longer motif. However, species-specific loci with the highest values of allelic richness (AH_051, AH_257, and AH_447) have a dinucleotide motif (Table 1 and 4). Consequently, because the heterozygosity is partly a function of the number of alleles (Zurn et al., 2020), we found a lower genetic diversity value for specific loci. However, F_{IS} and the frequency of null alleles were at a similar level (Tables 4 and S3).

Although a small number of individuals per taxon was used in our study, a simple test of Codom-Genotypic Genetic Distance between species performed in GENEALX (Peakall & Smouse, 2006) showed results consistent with the current systematic position of individual taxa (Fig. 2, Table S6). Species from section *Pavia* (*A. glabra*, *A. flava*, *A. ×neglecta*, and *A. pavia*) and *A. ×hybrida*, which is a hybrid between *A. flava* and *A. pavia*, were included in one group. The second group was formed by representatives of section *Aesculus* (*A. hippocastanum* and *A. turbinata*), section *Macrothyrus* (*A. parviflora*), and *A. ×carnea*, which is a hybrid between *A. hippocastanum* and *A. pavia*. A close relationship between sections *Aesculus* and *Macrothyrus* was previously reported in a phylogenetic study (Harris & Xiang, 2009). *Aesculus chinensis*, a representative of section *Calothyrsus*, was different from the other taxa which actually reflects its current systematic position (Harris & Xiang, 2009) (Fig. 2). In view of this, our results indicate that polymorphic microsatellites developed for one *Aesculus* species can be successfully used for other species from this genus in a situation of lack of the species-specific markers (Bačovský et al., 2017; Walas et al., 2019). In addition, we provide a list of nearly 500 pairs of primers that can be potentially applied in *A. hippocastanum* and other taxa (Table S1).

Analysis conducted in STRUCTURE, based on 20 loci, defined seven clusters of horse-chestnut, whereas for cross-amplified loci and species-specific loci it showed three and five groups, respectively. It is commonly known that an increased number of used markers may help in better recognition of the genetic structure and more correctly describes diversity indices. Inspection of barplots generated by STRUCTURE with species-specific and cross-amplified loci revealed some differences in population structure, especially in populations from the Pindos Mts. Based on species-specific SSRs, Kalampaka stand was defined as more distinct from the remaining populations from the northern Pindos Mts., i.e. Dasos Nanitsa and Ondria. The distinctiveness of Kalampaka may reflect the particular demographic

history of this genetically depleted population (Table 5). Both marker sets are convergent in this aspect and show low allelic and gene diversity of Kalampaka in contrast to other populations (Walas et al., 2019) that may suggest drift-induced differentiation detected by STRUCTURE analysis. Despite that, another population from this area, Perivoli, was constantly different from the neighbouring stands and showed genetic affinities to the marginally located population from Karitsa, irrespectively of the markers set (Fig. 3). Similarly, population Mariolata from the southern edge of the species range was always located in the distinct cluster using either specific or cross-amplified markers, which confirms that both marker sets are largely convergent in the detection of the genetic structure. The major discrepancy relates to the rate of the detected admixture which was higher for specific loci than for non-specific loci. Accordingly, Ondria was the most admixed population because only 30% of the individuals from that stand possessed the genome with 95% of the membership to a single cluster. On the contrary, most of the individuals from Perivoli and Karitsa reached this high level of membership (65.38% and 70.83%, respectively). The most probable explanation for this situation is the presence of a lower allelic diversity discovered with the set of the species-specific loci that in turn, resulted in a lower resolution and individual genealogy of each locus that may have interfered with the results. The genotypic error rate that can overestimate the level of the admixture was too low in our case to have such an impact (Reeves et al., 2016).

UPGMA clustering (Fig. 4) based on Nei genetic distances revealed the pattern which is partly congruent with the Bayesian inferences and demonstrated the distinctiveness of the marginal population Mariolata. It confirmed also the similarity between Ondria and Kalampaka populations. Additionally, Vaeoni seemed to be genetically closer to Dasos Nanitsa than to Karitsa or Perivoli, which was confirmed also by the value of the pairwise F_{ST} (Table S7) but disagrees with the STRUCTURE results. The main conclusion that can be drawn from the cluster analysis is the general genetic similarity of the populations from the Pindos Mts. except for Perivoli that was also detected by the Bayesian approach (Fig. 3). In pairwise comparisons, the highest genetic differentiation was noted between Kalampaka and Karitsa that originate from distinct mountain ranges (Table S7). Generally, the pairwise differentiation values were high which agrees with the Bayesian inferences about significant genetic structure in *A. hippocastanum*.

The genetic separation of the populations from the Pindos Mts., the Ossa Massif, and the Parnassus Massif as revealed by different methods used in this study, reflects spatial isolation of different mountain

ranges that currently provide suitable habitats for horse-chestnut. This pattern of differentiation is commonly reported for species inhabiting mountains (Noguerales et al., 2016). Horse-chestnut occurs in a topographically complex landscape, frequently in inaccessible mountainous habitats, which reduces the opportunity of gene flow at the landscape level. This particularly refers to Perivoli that despite its location in Pindos Mts. lacks significant similarity to other stands from this mountain range which likely demonstrates limited gene flow that was also suggested in the previous estimation of migration rate (Walas et al., 2019). Additionally, the reproduction mode and species biology limit genetic connectivity among populations separated geographically.

An efficient allocation of the conservation efforts in the case of the endemic and relic species requires broad knowledge not only about their ecology but also their genetic structure, its attributes, and factors. Based on such collection of the data, efficient Conservation Units may be set to protect not only the biological objects (populations and species) but also to support the maintenance of the evolutionary processes (Médail & Baumel, 2018). Microsatellite markers can be successfully used in conservation prioritization (Mattioni, 2017; Médail & Baumel, 2018). The reserve selection analysis, conducted in DIVA-GIS software (Hijmans et al., 2001), recognised Dasos Nanitsa and Karitsa as populations with the highest priority for conservation (Fig. 5). Unfortunately, Dasos Nanitsa, as well as other populations from the northern Pindos, are not recognised as the Sites of Community Importance (SCI) designed for the protection of flora and fauna in Greece. It is even more worrying as the horse-chestnut has not been included in the Red List of Greece (Phitos et al., 2009). Thus, a large part of the gene pool of this species may be lost as a result of a lack of formal protection. Populations from the edge of the species range, like Mariolata and Karitsa, which also harbour a high genetic diversity can be endangered with extinction because of the projected climatic changes (Walas et al., 2019). Their peripheral geographical location that implies isolation and ecological marginality as revealed in ecological niche modelling, may induce population reduction and decline (Walas et al., 2019; Schueler et al., 2014). Presented results are focused only on the Greek range while getting the deep insight into the species genetic resources would require the inclusion of populations from Albania and North Macedonia into the investigation.

Microsatellite analysis is easy to perform and is highly informative; therefore it is a widely used tool in population and conservation genetics studies (Madesis et al., 2013; Vieira et al., 2016). Although many new methods and approaches have appeared in molecular biology due to technological advancement,

SSRs remain useful, and even comparable with SNPs (Emanuelli et al., 2013; Filippi et al., 2015; Hodel et al., 2017; Zurn et al., 2020). If we think about a genetic tool that is easy and repeatedly used in routine analysis aiming to support conservationists and practitioners in their decisions, microsatellite markers fulfil these requirements in terms of the technological and biological aspects. Integration of genetic methods into conservation management, though it is still not a routine procedure (Taylor et al., 2017), becomes a more and more important approach that may facilitate the development of an effective strategy for the conservation of living organisms (McMahon et al., 2014).

Acknowledgements

This study was financed by the Polish National Science Centre (2017/27/N/NZ8/02781) and partly the statutory works of the Institute of Dendrology, Polish Academy of Sciences. We thank M. Łuczak for laboratory support.

References

- Allen DJ & Khela S (2017) *Aesculus hippocastanum* (errata version published in 2018). The IUCN Red List of Threatened Species 2017: e.T202914A122961065.
- Bačovský V, Vyhnánek T, Hanáček P, Mertelík J & Šafránková I (2017) Genetic diversity of chestnut tree in relation to susceptibility to leaf miner (*Cameraria ohridella* Deschka & Dimič). *Trees* 31: 753–763. doi:10.1007/s00468-016-1506-2.
- Boratyński A, Wachowiak W, Dering M, Boratyńska K, Sękiewicz K, Sobierajska K, Jasińska AK, Klimko M, Montserrat JM, Romo A, Ok T & Didukh Y (2014) The biogeography and genetic relationships of *Juniperus oxycedrus* and related taxa from the Mediterranean and Macaronesian regions. *Botanical Journal of the Linnean Society* 174: 637–653. doi:10.1111/boj.12147.
- Casazza G, Giordani P, Benesperi R, Foggi B, Viciani D, Filigheddu R, Farris E, Bagella S, Pisanu S & Mariotti MG (2014) Climate change hastens the urgency of conservation for range-restricted plant species in the central-northern Mediterranean region. *Biological Conservation* 179: 129–138. doi:10.1016/j.biocon.2014.09.015.
- Chapuis M-P & Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution* 24: 621–631. doi:10.1093/molbev/msl191.
- Chybicki I (2016) INEST 2.0 [Computer Software]. https://www.ukw.edu.pl/pracownicy/strona/igor_chybicki/software_ukw/.

- Dumolin S, Demesure B & Petit RJ (1995) Inheritance of chloroplast and mitochondrial genomes in pedunculate oak investigated with an efficient PCR method. *Theoretical and Applied Genetics* 91: 1253–1256. doi:10.1007/BF00220937.
- Emanuelli F, Lorenzi S, Grzeskowiak L, Catalano V, Stefanini M, Troggio M, Myles S, Martinez-Zapater JM, Zyprian E, Moreira FM & Grandó MS (2013) Genetic diversity and population structure assessed by SSR and SNP markers in a large germplasm collection of grape. *BMC Plant Biology* 13: 39. doi:10.1186/1471-2229-13-39.
- Filippi CV, Aguirre N, Rivas JG, Zubrzycki J, Puebla A, Cordes D, Moreno MV, Fusari CM, Alvarez D, Heinz RA, Hopp HE, Paniago NB & Lia VV (2015) Population structure and genetic diversity characterization of a sunflower association mapping population using SSR and SNP markers. *BMC Plant Biology* 15: 1–12. doi:10.1186/s12870-014-0360-x.
- Giorgi F & Lionello P (2008) Climate change projections for the Mediterranean region. *Global and Planetary Change* 63: 90–104. doi:10.1016/j.gloplacha.2007.09.005.
- Godoy FMR, Paggi GM & Palma-Silva C (2019) Cross-amplification of nuclear microsatellite markers in *Aechmea distichantha* Lem. (Bromeliaceae). *Brazilian Journal of Botany* 42: 353–359. doi:10.1007/s40415-019-00528-z.
- Goudet J (2001) FSTAT, a program to estimate and test gene diversity and fixation indices (version 2.9.3). <http://www2.unil.ch/popgen/softwares/fstat.htm>.
- Harris AJ & Xiang QY (2009) Estimating ancestral distributions of lineages with uncertain sister groups: a statistical approach to Dispersal–Vicariance Analysis and a case using *Aesculus* L. (Sapindaceae) including fossils. *Journal of Systematics and Evolution* 47: 349–368. doi:10.1111/j.1759-6831.2009.00044.x.
- Harrison S & Noss R (2017) Endemism hotspots are linked to stable climatic refugia. *Annals of Botany* 119: 207–214. doi:10.1093/aob/mcw248.
- Hijmans RJ, Guarino L, Cruz M & Rojas E (2001) Computer tools for spatial analysis of plant genetic resources data: 1. DIVA-GIS. *Plant Genetic Resources Newsletter* 127: 15–19.
- Hodel RGJ, Chen S, Payton AC, McDaniel SF, Soltis P & Soltis DE (2017) Adding loci improves phylogeographic resolution in red mangroves despite increased missing data: comparing microsatellites and RAD-Seq and investigating loci filtering. *Scientific Reports* 7: e17598. doi:10.1038/s41598-017-16810-7.
- Jagiello R, Baraniak E, Karolewski P, Łakomy P, Behnke-Borowczyk J, Walczak U & Giertych MJ (2017) Ecophysiological aspects of the interaction between *Cameraria ohridella* and *Guignardia aesculi* on *Aesculus hippocastanum*. *Dendrobiology* 78: 146–156. doi:10.12657/denbio.078.014.
- Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA & Mayrose I (2015) Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources* 15: 1179–1191. doi:10.1111/1755-0998.12387.
- Koressaar T & Remm M (2007) Enhancements and modifications of primer design program Primer3. *Bioinformatics* 23: 1289–1291. doi:10.1093/bioinformatics/btm091.
- Lack HW (2000) Lilac and horse-chestnut: discovery and rediscovery. *Curtis's Botanical Magazine* 17: 109–141. doi:10.1111/1467-8748.00255.
- Madesis P, Ganopoulos I & Tsaftaris A (2013) Microsatellites: evolution and contribution: Microsatellites (ed. by SK Kantartzi) Humana Press, Totowa, NJ, pp. 1–13. doi:10.1007/978-1-62703-389-3_1.
- Mattioni C, Martin MA, Chiocchini F, Cherubini M, Gaudet M, Pollegioni P, Velichkov I, Jarman R, Chambers FM, Paule L, Damian VL, Crainic GC & Villani F (2017) Landscape genetics structure of European sweet chestnut (*Castanea sativa* Mill): indications for conservation priorities. *Tree Genetics & Genomes* 13: e39. doi:10.1007/s11295-017-1123-2.
- McMahon BJ, Teeling EC & Höglund J (2014) How and why should we implement genomics into conservation? *Evolutionary Applications* 2014 7: 999–1007. doi:10.1111/eva.12193.
- Médail F & Baumel A (2018) Using phylogeography to define conservation priorities: The case of narrow endemic plants in the Mediterranean Basin hotspot. *Biological Conservation* 224: 258–266. doi:10.1016/j.biocon.2018.05.028.
- Minami E, Isagi Y, Kaneko Y & Kawaguchi (1998) Polymorphic microsatellite markers in Japanese horse chestnut *Aesculus turbinata* Blume. *Molecular Ecology* 7: 1616–1617.
- Mora F, Arriagada O, Ballesta P & Ruiz E (2017) Genetic diversity and population structure of a drought-tolerant species of *Eucalyptus*, using microsatellite markers. *Journal of Plant Biochemistry and Biotechnology* 26: 274–281. doi:10.1007/s13562-016-0389-z.
- Myers N, Mittermeier RA, Mittermeier CG, Da Fonseca GA & Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature* 403: 853–858. doi:10.1038/35002501.
- Noguerales V, Cordero PJ & Ortego J (2016) Hierarchical genetic structure shaped by topography in a narrow-endemic montane grasshopper. *BMC Evolutionary Biology* 16: 96. doi:10.1186/s12862-016-0663-7.

- Orsenigo S, Montagnani C, Fenu G, Gargano D, Peruzzi L, Abeli T, Alessandrini A, Bacchetta G, Bartolucci F, Bovio M, Brullo C, Brullo S, Carta A, Castello M, Cogoni D, Conti F, Domina G, Foggi B, Gennai M, Gigante D, Iberite M, Lasen C, Magrini S, Perrino EV, Prosser F, Santangelo A, Selvaggi A, Stinca A, Vagge I, Villani M, Wagensommer RP, Wilhalm T, Tartaglino N, Duprè E, Blasi C & Rossi G (2018) Red Listing plants under full national responsibility: Extinction risk and threats in the vascular flora endemic to Italy. *Biological Conservation* 224: 213–222. doi:10.1016/j.biocon.2018.05.030.
- Paradis E (2010) pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics* 26: 419–420. doi:10.1093/bioinformatics/btp696.
- Paradis E & Schliep K (2019) ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35: 526–528.
- Peakall ROD, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Resources* 6: 288–295. doi:10.1111/j.1471-8286.2005.01155.x.
- Pecl GT, Araújo MB, Bell JD, Blanchard J, Bonebrake TC, Chen I-C, Clark TD, Colwell RK, Danielsen F, Evengård B, Falconi L, Ferrier S, Frusher S, Garcia RA, Griffis RB, Hobday AJ, Janion-Scheepers C, Jarzyna MA, Jennings S, Lenoir J, Linnetved HI, Martin VY, McCormack PC, McDonald J, Mitchell NJ, Mustonen T, Pandolfi JM, Pettorelli N, Popova E, Robinson SA, Scheffers BR, Shaw JD, Sorte CJB, Strugnell JM, Sunday JM, Tuanmu M-N, Vergés A, Villanueva C, Wernberg T, Wapstra E & Williams SE (2017) Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. *Science* 355. doi:10.1126/science.aai9214.
- Peñuelas J, Sardans J, Estiarte M, Ogaya R, Carnicer J, Coll M, Barbeta A, Rivas-Ubach A, Llusà J, Garbulsky M (2013) Evidence of current impact of climate change on life: a walk from genes to the biosphere. *Global Change Biology* 19: 2303–2338. doi:10.1111/gcb.12143.
- Phitos D, Constantinidis T & Kamari G (2009) The red data book of rare and threatened plants of Greece. Hellenic Botanical Society, Patras.
- Pritchard JK, Stephens M & Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959. doi:10.1093/genetics/155.2.945.
- QGIS Development Team (2012) QGIS Geographic Information System. Open Source Geospatial Foundation Project. Available at: <http://qgis.osgeo.org>.
- R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: <http://www.R-project.org/>.
- Ravishankar KV, Mani BH-R, Anand L & Dinesh MR (2011) Development of new microsatellite markers from Mango (*Mangifera indica*) and cross-species amplification. *American Journal of Botany* 98: e96–99. doi:10.3732/ajb.1000263.
- Raymond M (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* 86: 248–249.
- Reeves PA, Bowker CL, Fetting CE, Tembrock LR & Richards CM (2016) Effect of error and missing data on population structure inference using microsatellite data. *bioRxiv* doi:10.1101/080630.
- Rousset F (2008) GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources* 8: 103–106. doi:10.1111/j.1471-8286.2007.01931.x.
- Sarkar D (2008) Lattice: Multivariate data visualization with R. Springer, New York. <http://lmdvr.r-forge.r-project.org>.
- Schueler S, Falk W, Koskela J, Lefèvre F, Bozzano M, Hubert J, Kraigher H, Longauer R & Olrik, DC (2014) Vulnerability of dynamic genetic conservation units of forest trees in Europe to climate change. *Global Change Biology* 20: 1498–1511. doi:10.1111/gcb.12476.
- Selkoe KA & Toonen RJ (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters* 9: 615–629. doi:10.1111/j.1461-0248.2006.00889.x.
- Steele H, Laue BE, MacAskill GA, Hendry SJ & Green S (2010) Analysis of the natural infection of European horse chestnut (*Aesculus hippocastanum*) by *Pseudomonas syringae* pv. *aesculi*. *Plant Pathology* 59: 1005–1013. doi:10.1111/j.1365-3059.2010.02354.x.
- Sugai K, Watanabe K, Kato H & Sugawara T (2016) Development of SSR markers for *Psychotria homalosperma* (Rubiaceae) and cross-amplification in four other species. *Applications in Plant Sciences* 4: 1500133. doi:10.3732/apps.1500133.
- Taylor HR, Dussex N & van Heezik Y (2017) Bridging the conservation genetics gap by identifying barriers to implementation for conservation practitioners. *Global Ecology and Conservation* 10: 231–242. doi:10.1016/j.gecco.2017.04.001.
- Thalman C, Freise J, Heitland W & Bacher S (2003) Effects of defoliation by horse chestnut leafminer (*Cameraria ohridella*) on reproduction in *Aesculus hippocastanum*. *Trees* 17: 383–388. doi:10.1007/s00468-003-0249-z.
- Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M & Rozen SG (2012) Prim-

- er3 – new capabilities and interfaces. *Nucleic Acids Research* 40: e115. doi:10.1093/nar/gks596.
- Vieira MLC, Santini L, Diniz AL & Munhoz C de F (2016) Microsatellite markers: what they mean and why they are so useful. *Genetics and Molecular Biology* 39: 312–328. doi:10.1590/1678-4685-GMB-2016-0027.
- Vignes H, Hossaert Mckey M, Beaune D, Fevre D, Anstett M-C, Borges RM, Kjellberg F & Chevallier MH (2006) Development and characterization of microsatellite markers for a monoecious *Ficus* species, *Ficus insipida*, and cross-species amplification among different sections of *Ficus*. *Molecular Ecology Resources* 6: 792–795. doi:10.1111/j.1471-8286.2006.01347.x.
- Walas Ł, Dering M, Ganatsas P, Pietras M, Pers-Kamczyc E & Iszkuło G (2018) The present status and potential distribution of relict populations of *Aesculus hippocastanum* L. in Greece and the diverse infestation by *Cameraria ohridella* Deschka & Dimić. *Plant Biosystems* 152: 1048–1058. doi:10.1080/11263504.2017.1415991.
- Walas Ł, Ganatsas P, Iszkuło G, Thomas PA & Dering M (2019) Spatial genetic structure and diversity of natural populations of *Aesculus hippocastanum* L. in Greece. *PloS One* 14: e0226225. doi:10.1371/journal.pone.0226225.
- Wang J (2018) Estimating genotyping errors from genotype and reconstructed pedigree data. *Methods in Ecology and Evolution* 9: 109–120. doi:10.1111/2041-210X.12859.
- Zurn JD, Nyberg A, Montanari S, Postman J, Neale D & Bassil N (2020) A new SSR fingerprinting set and its comparison to existing SSR-and SNP-based genotyping platforms to manage *Pyrus* germplasm resources. *Tree Genetics & Genomes* 16: 1–10. doi:10.1007/s11295-020-01467-7.