Tree Genetics & Genomes (2017) 13:52
DOI 10.1007/s11295-017-1134-z

ORIGINAL ARTICLE

To what extent do wild apples in Kazakhstan retain their genetic integrity?

Madina Y. Omasheva[1] · Henryk Flachowsky[2] · Natalya A. Ryabushkina[1] · Alexandr S. Pozharskiy[1] · Nurbol N. Galiakparov[1] · Magda-Viola Hanke[2]

Received: 7 June 2016 / Revised: 14 March 2017 / Accepted: 20 March 2017

Abstract Kazakhstan belongs to the center of origin of apple. *Malus sieversii* (Ledeb.) M. Roem., the ancestral progenitor of the cultivated apple is native to this region. Pressure on the natural habitats of this wild apple has been intensified due to agriculture, grazing, and urbanization in the last century. For decades, *M. sieversii* in Kazakhstan has been subjected to the "Red Book of the Kazakh SSR" and today, this species is threatened with extinction. Wild apple undergoes exceptional losses in habitats, and the risk for losing the genetic integrity becomes worse due to increasing cultivation of cultivated apples and frequently occurring crosspollination events. The present study was focused on the current state of *M. sieversii* in Kazakhstan, the level of its diversity, its genetic integrity, and the identification of regions where future activities for conservation will have a good chance of success. A total of 311 *M. sieversii* samples of 12 populations collected in the wild, 16 previously selected wild apple genotypes, and 50 grown cultivars were studied using 16 simple sequence repeat (SSR) markers for genetic analysis. The results suggest that the level of genetic diversity is high. The differentiation between the populations was low, although the within-population heterozygosity was relatively high. A significant number of hybrids (8–95%) between *M. sieversii* and cultivated apples were found suggesting frequent crop-to-wild gene flow. The percentage of pure wild apple genotypes was highest in Krutoe truct and Tauturgen. These sites should be taken into account for future in situ long-term preservation activities.

Keywords *Malus sieversii* · Microsatellite markers · Genetic diversity · Population structure

Communicated by E. Dirlewanger

Electronic supplementary material The online version of this article (doi:10.1007/s11295-017-1134-z) contains supplementary material, which is available to authorized users.

✉ Magda-Viola Hanke
viola.hanke@julius-kuehn.de

[1] Institute of Plant Biology and Biotechnology, Timiriazev 45, Almaty, Kazakhstan 05040

[2] Julius Kühn-Institute, Federal Research Institute for Cultivated Plants, Institute for Breeding Research on Fruit Crops, Pillnitzer Platz 3a, 01326 Dresden, Germany

Introduction

Kazakhstan covers the northern region of the South Western Asian Centre of origin according to N. Vavilov where a number of plants, like the cultivated apple (*Malus domestica* Borkh.), were domesticated (Vavilov 1931). *Malus sieversii* (Ledeb.) M. Roem., the ancestral progenitor of modern apple cultivars, is native to this region. Natural habitats of this wild apple species can be found in Kazakhstan, Kyrgyzstan, Tajikistan, Uzbekistan, and the western part of China (Harris et al. 2002; Velasco et al. 2010; Cornille et al. 2014). *M. sieversii* was first discovered in Kazakhstan in 1796 by Johann August Carl Sievers in the Tarbagatai Mountains and later described by the German botanist Carl F. von Ledebour in 1830 (Dzhangaliev 1977). Trees of *M. sieversii* are 2 to 10 m (sometimes 14 m) tall with yellowish green fruits tinged with red. Fruits of *M. sieversii* have a globose or depressed globose shape and a size of 3 to 4.5 cm (sometimes up to 7 cm) in diameter (Cuizhi and Spongberg 2003). Its close genetic relationship to the cultivated apple (Harris et al. 2002; Velasco et al. 2010), together with some special traits like fruit size, taste, resistance to apple scab, fire blight, drought, and numerous soil pathogens (Ignatov and Bodishevskaya 2011; Pons 2006) make this species interesting for fruit breeders and botanists.

Published online: 07 April 2017

Although this species was neglected in the past, it has always contributed significantly to human's nutrition. Towards the end of the twentieth century and the beginning of the twenty-first century, the international scientific community started to pay more attention to *M. sieversii*. This attention was a result of the activities of the Kazakh scientist, Aimak D. Dzhangaliev, who carried out inventory of this species in its natural habitats. He also described numerous apple trees growing naturally in the mountains on northern, north-eastern, and north-western slopes at 800 to 1500 m, and in some locations at up to 2000 m above sea level. He discovered trees of *M. sieversii* mainly within the low forest belt, in forest meadows, in forest steppes, and in steppe belts of deciduous forest associations. Most of these trees were grown at forest borders and in shrub communities along the mountainous ridges extending from eastern through south-eastern to southern Kazakhstan (Dzhangaliev 1977, 2003).

Furthermore, Aimak D. Dzhangaliev described the high level of intraspecific diversity of *M. sieversii* connected with mountain regions differing in their geological history, their geographical structure, and their natural conditions. Based on observations in the twentieth century, a number of different mountain floristic regions were identified in Kazakhstan where wild apples occur naturally (Turehanova 2012).

However, the high level of variation between individual trees of wild-grown apples in Kazakhstan has led to several debates regarding the exact number of different species in the past. Six different apple species were described by Bykov (1961), whereas other authors suggested two or three species with unclear species boundaries. Dzhangaliev defined the number of species as three and assigned all individuals (with exception of a few aliens) to *Malus kirghisorum*, *Malus niedzwetzkyana*, and *M. sieversii* as the most prominent one in Kazakhstan (Dzhangaliev 1977, 2003). In 2009, Gayle M. Volk assigned wild apples collected in Kyrgyzstan to a common genetic lineage with *M. sieversii* individuals from the Karatau Mountain range of Kazakhstan (Volk et al. 2009). Based on genetic fingerprint information, it became obvious that *M. sieversii* is a highly diverse species with a range in genetic and phenotypic trait variations (Volk et al. 2013).

Pressure on the wild apple forests in Kazakhstan has intensified for many decades due to urbanization, agriculture, grazing, and wood harvesting. In the late nineteenth century, many wild fruit trees in the foothills of Zailiysky Alatau were grubbed by local people and replaced by agricultural crops (Severtsov 1873). Between 1932 and 1967, wild apple trees were often used as rootstocks and cultivated apples were grafted on their top in the forest. As a consequence, the natural forest orchard systems lost species-specific dynamic features and genetic integrity. Existing wood stands of apple were not destroyed permanently, but their ability for natural renewal was significantly reduced (Dzhangaliev 2003). In 1981, *M. sieversii* and *M. niedzwetzkyana* were included in the "Red Book of Kazakh SSR." Nevertheless, the progressive reduction in the number of wild apple representatives continued and as a result, the natural habitat of *M. sieversii* has declined by over 70% in Kazakhstan during the last 30 years (Eastwood et al. 2009). This bad situation has been worsened by the adoption of the new Land Code in 2003 which approves the private property on land. Nowadays, *M. sieversii* is a species with continuously declining number of individuals (Ivashenko 2005) thus threatened with extinction in its natural environment (Eastwood et al. 2009).

The work presented here was initiated to study the current status of Kazakhstan wild apple populations. Three hundred eleven individual trees of 12 natural *M. sieversii* populations belonging to three geographical areas were evaluated on their genetic diversity and integrity using a set of 16 simple sequence repeat (SSR) markers. The populations were collected in a geographical range covering the eastern, south-eastern, and southern mountainous regions of Kazakhstan. A genetic population structure analysis revealed the presence of genetic structures across all populations regardless of existing distances between their geographical locations, but also a high level of admixture with *M. domestica* in some populations. The relatively high number of hybrids in some populations, which results most probably from crosspollination events between *M. sieversii* and *M. domestica* in regions of spatial proximity between both species, underscores the alarming situation of the ongoing fragmentation of wild species populations in their natural habitats.

Materials and methods

Plant material

Leaf materials of 311 *M. sieversii* trees were collected at 12 different sites assorted as three different geographical groups in Kazakhstan (Fig. 1, Table 1). In large populations (≥30 individuals), 25–40 individuals were selected at random with regards to the age of the trees. The approximate age of the trees was determined using a conventional method used in forest taxation (Forest encyclopedia 1986). In each region, the mean value of annual increments of randomly selected trees was estimated and then the age was determined by the diameter of tree trunk at a height of 1.3 m above soil level. In small populations, mostly all individuals were selected. In addition, 16 apple clones were collected in Tarbagatay (five clones), Dzhungarskiy Alatau (nine clones), and Zailiysky Alatau (two clones) and included in this study. These clones were selected in 1990 by A. D. Dzhangaliev based on their superior fruit size and other parameters with interest for breeding. Finally, 50 apple cultivars being grown in Kazakhstan were included as references to estimate the level of admixture.

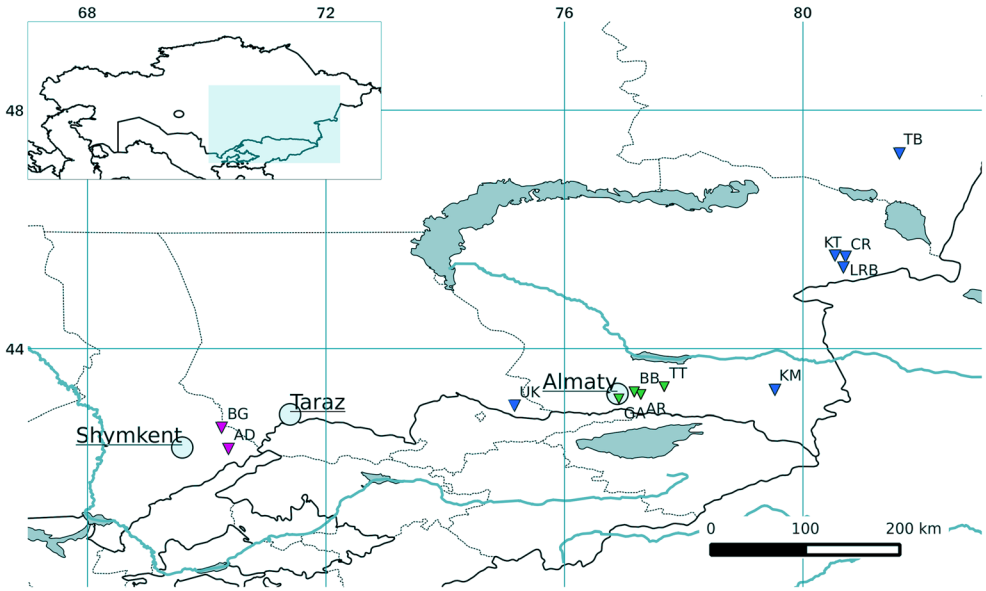

Fig. 1 Map of Kazakhstan with geographical distribution of the 12 sites for collecting *Malus sieversii* samples examined in this study. Sites of *M. sieversii* populations were marked by differently *colored triangles*. All populations marked with *blue triangles* belong to the geographical region I. Populations marked with *green triangles* belong to the geographical region II, whereas populations with *pink triangles* belong to III. *TB* Tarbagatay, *CR* Chernoff River, *KT* Krutoe truct, *LRB* Lepsy right bank, *KM* Ketmen, *UK* Uryukty, *TT* Tauturgen, *AR* Almaty reserve, *BB* Belbulak, *GA* Great Almaty gorge, *BG* Bozturgay gorge, *AD* Aksu Dzhabagly

Genomic DNA extraction

Genomic DNA was extracted from 100 mg of fresh leaf tissue as described by Aubakirova et al. (2014). Quality was tested by 1% agarose gel electrophoresis and the NanoDrop 2000 (Thermo Scientific, USA) was used for calculating the concentration. DNA was diluted to 30 ng/μL.

SSR marker analysis

Sixteen SSR markers namely GD12, GD96, GD142, GD147, GD162, GD100, CH01h10, CH01h01, CH04c07, Hi02c07, CH01f03b, CH02d08, CH02c11, CH04e05, CH01f02 and CH02c09 (Hokanson et al. 1998; Hemmat et al. 2003; Liebhard et al. 2002; Richards et al. 2009a) were used. These markers, which are distributed across the apple genome, were chosen due to their use in similar studies (e.g., Richards et al. 2009a) and their belonging to the marker set suggested for genotyping of apple genetic resources by the European Cooperative Programme for Plant Genetic Resources (ECPGR). For example, CH04e05, CH02c11, CH02c09, CH02d08, CH04c07, CH01h01, Hi02c07, and CH01h10 belong to the priority group 1 of the ECPGR marker set whereas CH01f02, CH01f03b, GD12, and GD147 belong to priority group 2. Markers GD12, GD96, GD142, GD147, GD162, and GD100 were used in combination with three universal tail oligos D8S1132 (VIC), D12S1090 (NED), and DYS437 (FAM) as described by Missiaggia and Grattapaglia (2006). PCR was performed using a Mastercycler Pro S thermocycler (Eppendorf, Hamburg, Germany). A first-step amplification was carried out in a volume of 20 μl, containing 20 ng genomic DNA, 1× Taq buffer (750 mM Tris HCl, pH 8.8, 200 mM $(NH_4)_2SO_4$, 0.1% Tween 20), 2.5 mM $MgCl_2$, 0.2 mM dNTPs, 0.2 mM of each of the respective SSR primers, and 1 unit Tag polymerase (Thermo Scientific, USA). After denaturation at 94 °C for 2 min, seven PCR cycles were performed with 1 min at 94 °C, 2 min at 60 °C, and 2 min at 72 °C. Subsequently, 20 cycles with 1 min at 94 °C, 2 min at 54 °C, 2 min at 72 °C, and final extension of 10 min at 72 °C were performed. Five microliters of the PCR products were checked on a 1% agarose gel. The remaining products were tenfold diluted with ddH_2O. One-microliter diluted PCR product was used in a second PCR where the tail primers were used instead of the forward primers. The same PCR program described previously was used. After a final extension of 30 min, the products of all six standard markers per sample were combined and mixed with formamide and size standard 500 LIZ (Applied Biosystems, Foster City, USA). Hence, in a total volume of 10 μl, fluorescent dye VIC was diluted 540-fold, NED 120-fold, and FAM 360-fold. Samples were analyzed on a capillary sequencer ABI Prizm 310 (Applied Biosystems, Foster City, CA, USA). Data were processed by GeneMapper Software 4.0 (Applied Biosystems, Foster City, CA, USA). The Type-it® Microsatellite PCR Kit (Qiagen GmbH, Hilden, Germany) was used for PCR reactions on markers CH01h10, CH01h01, CH04c07, Hi02c07, CH01f03b, CH02d08, CH02c11, CH04e05, CH01f02, and CH02c09. Markers were divided into four multiplex groups and each PCR reaction contained 10 ng genomic DNA, 1× PCR Master Mix, 0.5 μM of each primer and 0.5× Q-Solution in

Table 1 Physical description of the 12 sites in Kazakhstan, where samples of *M. sieversii* were collected

Geographical group	Region	FR	Sampling site (population)	Latitude	Longitude	Elevation (m)	Habitat	Number	Number of trees of different age		
									Up to 20 years	20–50 years	Older than 50 years
I	Tarbagatay	23	Tarbagatay (TB)	47° 16,830	81° 36,150	1015	Shrub meadow on a slope, south-eastern exposure	34	23	9	2
	Dzhungarskyi Alatau	24	Chernoff river (CR)	45° 31,275	80° 42,830	1240	Apple grove on the southern slopes and terraces	30	6	0	24
			Krutoe truct (KT)	45° 33,130	80° 43,840	1515	Forest-meadow border on the slope of the watershed, western exposure	28	8	15	5
			Lepsy right bank (LRB)	45° 32,955	80° 41,830	1160	Apple grove on the western slope alongside a pasture on the hillside	22	4	17	1
	Great Kirgizsay	25a	Ketmen (KM)	43° 19,195	79° 30,900	1440	Single apple trees on the bottom of the gorge	7	0	4	3
	Chu-Ili Mountains	26	Uryukty (UK)	43° 02,318	75° 09,637	1240	Single apple trees among small trees and shrubs (mini-ecosystem tugai)	6	0	4	2
II	Zailiysky Alatau	25	Tauturgen (Kuznetsov cleft) (TT)	43° 21,460	77° 40,355	1585	Apple forest-meadow on the northern slopes	40	10	11	19
			Almaty reserve (right Talgar) (AR)	43° 13,720	77° 16,880	1560	Apple grove on the terrace slope, north-eastern exposure	48	18	20	10
			Belbulak (BB)	43° 16,275	77° 10,255	1255	Northern exposure of the forest slope	30	5	6	19
			Great Almaty gorge (GA)	43° 09,135	76° 54,690	1320	Apple grove on the northern slope exposure	30	0	7	23
III	Karatau	28	Bozturgay gorge (BG)	42° 40,570	70° 15,630	845	At the floodplain among small trees and shrubs (mini-ecosystem tugai)	17	10	7	0
	Talas Alatau	29	Aksu Dzhabagly (AD)	42° 19,545	70° 22,270	1365	Apple grove on the northern exposure slope	19	8	10	1

Collection sites of this study were in accordance to Dzhangaliev 1977)

FR floristic region, *N* sample size

a total volume of 10 μl. After an initial denaturation at 95 °C for 5 min, 28 cycles with 1 min at 95 °C, 1.5 min at 60 °C, and 0.5 min at 72 °C were performed using the same PCR machine as mentioned before. After a final extension at 60 °C for 30 min, the PCR products were diluted in a 1:20 or 1:30 ratio with ddH_2O depending on the multiplex type. Samples for separation consisted of 24.9 μl GenomeLab™ Sample Loading Solution, 0.1 μl GenomeLab™ DNA size standard-400, 1 μl diluted PCR product, and a drop of mineral oil. The GenomeLab™ GeXP software (Beckman Coulter GmbH, Krefeld, Germany) was used to calculate size of PCR fragments.

Statistical analysis

Conversion factor for each SSR marker was estimated with the TANDEM v1.09 software program (Matschiner and Salzburger 2009). The numbers of different alleles (N_a) and their frequency, the number of effective alleles (N_e), number of private alleles (N_o), expected and observed heterozygosity (H_e; H_o), the inbreeding coefficient within individuals relative to the subpopulation (F_{IS}), the inbreeding coefficient within subpopulations relative to the total as a measure for genetic differentiation (F_{ST}), the Shannon-Weaver Index of ecology (I), and Nei's genetic distance were calculated with GenAlEx 6.5 (Peakall and Smouse 2012). Polymorphism information content (PIC) was estimated according to Botstein et al. (1980) using the Excel Microsatellite Toolkit. Allelic richness (Rs) and private allele richness (PR) were calculated by the rarefaction method with HP-RARE software (Kalinowski 2005).

Cluster analysis, genetic distance analysis, and principal coordinates analysis (PCoA) were performed using R ver. 3.2.2 (R Development Core Team 2008) with additional packages "geosphere", "ecodist" (Goslee and Urban 2007), "poppr" (Kamvar et al. 2015), "ggplot2" (Wickham 2009), "adegenet" (Jombart 2008), "ade4" (Dray and Dufour 2007) (dependencies for "poppr"), and "ape" (Paradis et al. 2004). Mantel test was conducted in order to evaluate correlation of genetic and geographic distances between populations. A geographic distance matrix was calculated using coordinate data from Table 1 by geosphere::distm function. Mantel test was performed using ecodist::mantel function with 10,000 permutations (the level of confidence limits was set as 0.95; confidence limits were estimated with 10,000 iterations of bootstrap).

Population structure and hybridization analysis

For genetic structure analysis, a Bayesian model-based approach was used as implemented in STRUCTURE 2.3.4. Software (Pritchard et al. 2000). The admixture model was used as recommended by Hubisz et al. (2009). The burning period and MCMC were consisted of 100,000 iterations each. K values in a range from 2 to 10 were tested in 20 independent runs, and the method by Evanno et al. implemented in STRUCTURE HARVESTER was used to find the most appropriate K value (Evanno et al. 2005; Earl and von Holdt 2012). Subsequently, 100 runs for each K from 2 to 6 were performed and aligned by CLUMPP 1.2.2. (Jakobsson and Noah 2007). DISTRUCT 1.1 was used for visualization of the results (Rosenberg 2004). Individuals with a membership coefficient to the wild gene pool larger than 0.9 were considered as pure wild enabling to estimate the proportion of pure wild individuals in each population (Cornille et al. 2015).

Results

Overall frequencies of the alleles detected using SSR markers

A set of 16 SSR markers was used to evaluate 311 wild apple samples of 12 natural *M. sieversii* populations and 50 selected cultivars of *M. domestica*. All markers were found to be polymorphic with 9 to 24 or 9 to 18 alleles per locus for wild apples and cultivars, respectively (Table S1). The PIC values of the markers were always >0.6, except for marker CH02d08. This marker showed a lower PIC value (0.37) for the wild apples. Thus, all markers were informative as also shown by the Shannon's index (I) ranging from 0.86 for marker CH02d08 to 2.45 for marker GD142. The 16 SSR markers amplified a total of 259 alleles for 311 wild apples, and 203 alleles for 50 cultivars. CH01f02 was the most variable marker with 24 and 18 alleles and CH02d08 was the least variable one.

Wild apples represent a unique genetic diversity

The 311 wild apple samples represent a unique genetic diversity. Each individual could be assigned to a unique genotype. No duplicates were found, which could have originated from clonal propagation via root suckers. Triploids were also not detected.

Overall genetic diversity

The total number of alleles (N_a) was noticeably higher in comparison to the number of effective alleles (N_e) leading to the conclusion that only a few alleles contributed to the diversity. This was the case for wild apples, but also for cultivars. The values for H_o and H_e were relatively high suggesting a good level of diversity within the collected plant material.

Genetic diversity within the populations

Mean numbers of 16.2 and 12.7 alleles per locus and 6.3 and 9.1 alleles per population were found for the wild apples and cultivars, respectively (Tables S1 and S2). The highest (9.12) and lowest (5.0) mean number of alleles per locus was found for cultivars followed by the Bozturgay gorge (BG) and Aksu Dzhabagly (AD) wild apple populations, respectively (Table S2). The mean number of effective alleles per locus ranged between 5.7 for the BG population and 2.3 for the AD population. The private allelic richness was maximal in cultivars and minimal in the Tauturgen (TT) population (Table S2). For most of the populations, H_o was slightly less than H_e except for the cultivars and the AD apple population. The inbreeding coefficient (F_{IS}) was always low. The highest F_{IS} (0.1) was detected in the Almaty reserve (AR) population. F_{IS} was slightly negative in the cultivars and in the AD population. The Ketmen (KM) and Uryukty (UK) populations were not taken into account because of their small size.

Allelic richness

Allelic richness was analyzed per locus and population for standardized number of individuals in each population (Table 2). The highest allelic richness (13.23) was found with marker CH04e05 for the GA population. Among the wild apple populations, the highest values were detected for the BG population at the loci CH01f02, GD96, GD142, and GD147. The lowest allelic richness was in the AD population at 9 of the 16 loci with a minimal mean per site of 4.8. Although the mean allelic richness was comparable for the geographical groups I, II, and III (Table 2), there were differences at individual loci. In *M. domestica* compared to the groups, allelic richness at locus CH01f02 was maximal with 16.18 and minimal at locus CH01h10 with 7.71.

Population structure analysis

STRUCTURE analysis revealed the presence of genetic structures across populations of the three geographical groups (I, II, and III). The most likely number of clusters (K) was evaluated considering the plateau criterion (Fig. 2a) and the ΔK method with the highest ΔK value observed at $K = 2$ (Fig. 2b). The estimated population structure inferred from this analysis is shown in Fig. 2c. The highest ΔK value ($K = 2$) reflects the presence of two clusters in the inferred population structure analysis. Each cluster clearly corresponds to *M. sieversii* (green cluster) and *M. domestica* (red cluster) representatives, respectively (while not taking clones selected by Dzhangaliev in natural populations into account).

Populations located in the floristic regions 23 and 24 (eastern Kazakhstan, Table 1) were not separated from the TT and AR populations of the south-eastern mountain system of Zaylisky Alatau. However, there was high admixture with *M. domestica* in Lepsy right bank (LRB) as the most eastern population and Belbulak (BB), Great Almaty gorge (GA), and BG as the south-eastern populations. The red cluster included *M. domestica*, the AD population, and Dzhangaliev's apple clones with some admixture with *M. sieversii*. At $K = 3$, populations located in the floristic regions 23 and 24 were clearly separated (blue cluster) from the populations of the floristic regions 25 (green cluster). AD was admixed. At $K = 4$, the AD population formed its own genetic pool, whereas the UK, KM, and BG populations were admixed. At $K = 5$, 6, and more, there were no new clusters formed.

All samples of *M. domestica* were classified correctly and did not show admixture at any K checked, except the cultivar "Niedzwiecki" (MD33 Fig. S2). Niedzwiecki is known as a hybrid of *M. sieversii* f. *niedzwetzkyana* and *M. domestica*. Its hybrid nature is clearly supported by the STRUCTURE analysis.

Genetic distance analysis among populations

To evaluate genetic distance, genetic differentiation and structuring among populations the Nei's minimum distance and pairwise F_{ST} were calculated (Table 3). Relationships among populations were illustrated by a dendrogram and a PCoA plot (Fig. 3). Based on Fig. 3a and Table 3, the populations of the geographical group I collected in Krutoe truct (KT), Chernoff River (CR), LRB, and Tarbagatay (TB) were very close to each other. The same was found for the populations (GA, BB, AR, and TT) of the geographical group II. According to Nei's algorithm, the genetic distance between groups I and II was 0.24, between II and III 0.4, and between I and III 0.59. The genetic distance was positively correlated with the geographic distance. It was also confirmed by Mantel test, that showed a significant positive correlation of $R = 0.57$ ($p < 0.01$).

F_{ST} ranged in group I from 0.02 (CR/LRB) to 0.06 (KT/LRB). In group II, F_{ST} values ranged between 0.01 (BB/GA) and 0.04 (TT/GA), and in group III, it was 0.15. Among groups I and II, F_{ST} was 0.07; among II and III, 0.10; and among I and III, 0.15.

The two-dimensional PCoA plot (Fig. 3b) separated the 311 samples into two distinct clusters with the eastern group (geographical group I) and the south-eastern group (geographical group II). The third gene pool of the southern regions (geographical group III) was less uniform. The possible cause is the smaller number of samples.

Admixture with *M. domestica*

A unweighted pair group method with arithmetic mean (UPGMA) dendrogram based on Nei's genetic distances was calculated from the dataset of 16 SSRs across the 311

Table 2 Allelic richness

Site[a]	TB	CR	KT	LRB	TT	AR	BB	GA	BG	AD	Groups[b]			*M.* × *domestica*
Locus	I				II				III		I	II	III	
CH01h10	3.75	7.26	4.49	6.32	5.09	7.05	5.94	5.05	8.00	5.57	7.03	7.74	9.86	7.71
CH01h01	7.60	7.63	5.38	6.63	6.58	7.85	7.73	8.46	6.00	4.77	8.90	10.1	6.99	8.31
CH04c07	5.25	6.83	2.61	2.95	4.10	6.79	7.22	5.45	8.00	7.46	6.96	8.041	8.94	9.61
Hi02c07	5.51	5.95	5.04	5.94	5.65	6.12	7.72	6.46	7.00	4.78	6.53	8.18	7.94	9.41
CH01f03b	5.63	7.91	8.88	8.44	5.65	7.09	8.90	7.16	9.00	2.89	8.89	9.34	8.94	11.4
CH02d08	2.25	4.87	1.61	5.27	3.87	4.04	4.38	4.13	6.00	3.68	5.06	4.91	5.97	8.67
CH02c11	8.39	7.30	4.55	7.32	5.78	6.59	8.44	8.44	10.0	4.68	9.33	9.11	9.99	12.4
CH04e05	10.6	7.64	5.82	7.76	8.94	10.5	9.19	13.2	11.0	3.89	11.7	13.9	10.9	13.9
CH01f02	6.01	9.42	8.32	10.3	7.95	8.60	10.9	9.35	12.0	5.68	12.0	12.1	11.9	16.2
CH02c09	6.42	8.51	6.59	6.95	5.42	8.45	7.35	8.35	7.00	4.79	8.27	9.54	7.00	10.9
GD12	6.00	6.30	6.08	7.61	7.66	8.23	6.89	7.33	9.00	3.68	8.45	9.44	10.8	12.1
GD96	6.58	7.22	4.58	6.75	8.49	9.93	8.30	8.51	12.0	2.94	8.23	12.2	13.0	14.4
GD142	9.16	8.12	6.40	9.26	8.50	11.12	10.3	10.1	12.0	6.47	10.3	13.2	12.9	12.9
GD147	8.97	7.69	5.58	8.44	6.06	7.43	7.30	8.21	12.0	5.79	9.66	9.17	13.9	11.0
GD162	6.29	9.03	6.56	10.9	7.19	11.2	10.4	11.0	6.00	5.68	11.2	13.6	7.89	13.2
GD100	3.75	4.70	2.98	4.99	3.00	5.06	3.59	5.16	4.00	3.98	4.82	5.49	4.99	10.6
mean/site	6.39	7.27	5.34	7.24	6.28	7.88	7.79	7.90	8.69	4.80	8.59	9.77	9.51	11.4
N	34	30	28	22	40	48	30	30	17	19	114	148	36	50

Allelic richness was measured per locus and site (upper part of the table) and per locus and cluster (lower part of the table)

TB Tarbagatay, *CR* Chernoff River, *KT* Krutoe truct, *LRB* Lepsy right bank, *TT* Tauturgen, *AR* Almaty reserve, *BB* Belbulak, *GA* Great Almaty gorge, *BG* Bozturgay gorge, *AD* Aksu Dzhabagly

[a] Allelic richness standardized to 17 diploid individuals (34 genes)

[b] Allelic richness standardized to 35 diploid individuals (70 genes)

wild apple genotypes, 50 cultivars, and 16 clones selected by Dzhangaliev. The dendrogram is composed of two main clusters (Fig. S1). Cluster 1 (samples MD35 to MD01) is composed of 35 samples of *M. domestica* and the Dzhangaliev clones DZh03, DZh05, DZh14, and DZh15. Cluster 2 is divided into three sub-clusters. Sub-cluster 2.1 (samples MD32 to CR27) is composed of seven samples of *M. domestica*, six Dzhangaliev clones, and 19 samples of *M. sieversii*. Sub-cluster 2.2 (samples Dzh08 to KT26) contains four Dzhangaliev clones and three samples of *M. sieversii*. Sub-cluster 2.3 is further divided into two sub-sub-clusters. Sub-sub-cluster 2.3.1 (samples GA20 to AR45) are three samples of *M. sieversii* and the *M. domestica* sample MD38. Sub-sub-cluster 2.3.2 is divided into two sub-sub-sub-clusters with sub-sub-sub-cluster 2.3.2.1 (samples MD18 to AD12) containing the remaining six samples of *M. domestica* (except of MD33), two Dzhangaliev clones and 58 samples of *M. sieversii*. Sub-sub-sub-cluster 2.3.2.2 (sample KM04 to KT02) is the largest cluster. It contains all the remaining samples of *M. sieversii*. In this cluster, there is no Dzhangaliev clone and no sample of *M. domestica* grouped, except of the cultivar Niedzwiecki (sample MD33).

Several individuals of the *M. sieversii* populations showed a variable level of admixture with *M. domestica* (Fig. 2c, S2) ranging from nearly no admixture to almost 100%. Nearly no admixture (membership coefficient >0.9) was found in the TT (92% pure wild apples) and KT (89% pure wild apples) populations located at altitudes of 1585 and 1515 m above sea level, respectively. The percentages of pure wild genotypes was slightly lower (83 and 85%) in the south-eastern AR and the eastern TB populations located at altitudes of 1560 and 1015 m, respectively. In other south-eastern populations admixture was much higher. For example, the percentage of pure *M. sieversii* individuals was about 40% in the GA (altitude 1320 m) and 47% in the BB (altitude 1255 m) populations, but nearly all individuals of the BG and AD populations (southern site) were admixed. No correlation was found between the altitude and the level of admixture, but there was a tendency for admixture increasing from the eastern to the southern sites. No correlation was found between the level of admixture and the mean tree age per population. For example, in TB (two third of the trees aged up to 20 years) 85%, in CR (four fifths aged over 50 years) 67%, in LRB (four fifths aged 20–50 years) 36%, and in TT (~50% aged over 50 years) 92% of the individuals were pure *M. sieversii* genotypes. Too

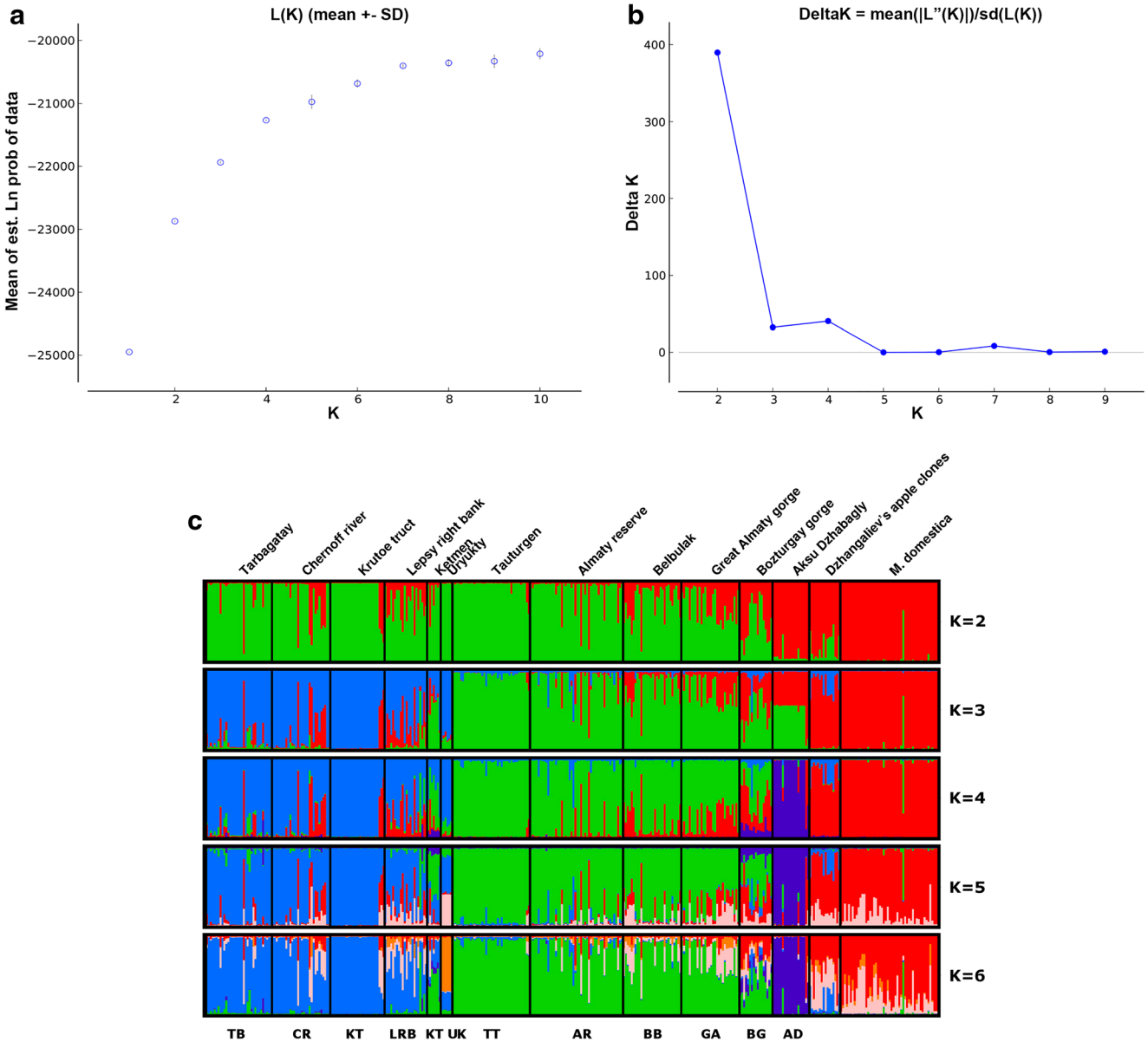

Fig. 2 Genetic structure analysis using 311 samples of *M. sieversii*, 16 clones selected by Dzhangaliev, and 50 cultivars of *M. domestica*. **a** Estimation of likelihood of STRUCTURE. The mean values of the logarithmic likelihood function $L(K)$ and standard deviation from 20 runs for each value of K (number of clusters) = 1–10. **b** The distribution of ΔK over K = 1–9 with ΔK as a ratio of the mean of $[L(K+1) - 2(L(K) + L(K-1)]$ to the standard deviation of $L(K)$. **c** Population structure inference by Bayesian assignment with K = 2–6. Each individual is represented by a *vertical line*. Populations are separated by a *vertical black line*. *Different colors in the same line* indicate the individual's admixture proportion (Q value) in K clusters

small populations (KM and UK) were not taken into account. Sixteen clones selected in wild populations by Dzhangaliev showed different levels of admixture with *M. domestica* (Fig. 1S, 2S).

Discussion

Malus sieversii in natural habitats is increasingly threatened by forest destruction, with its populations being restricted to areas that have been rapidly decreasing in size over the last decades (Zhang et al. 2007). Conservation of wild plant populations in their natural habitats is becoming a major concern. It requires detailed information about the occurrence and distribution of individual species, their genetic diversity, differentiation, and level of admixture with other species. The knowledge about these parameters is required for establishing sustainable management and conservation strategies (Cornille et al. 2015). The crabapple - *M. sieversii* (Ledeb.) M. Roem., is of particular interest as it is considered the main progenitor

Table 3 Genetic distance analysis using Nei's (1972) minimum distance (top diagonal) and pairwise F_{st} comparisons (bottom diagonal) among 12 populations of *M. sieversii*

	Population	FR	TB	CR	KT	LRB	TT	AR	BB	GA	BG	AD
I	TB	23	***	0.11	0.11	0.11	0.28	0.27	0.30	0.35	0.46	0.88
	CR	24	0.03	***	0.09	0.11	0.34	0.29	0.32	0.38	0.46	0.80
	KT	24	0.05	0.04	***	0.14	0.37	0.38	0.40	0.47	0.57	1.03
	LRB	24	0.02	0.02	0.06	***	0.30	0.26	0.21	0.28	0.37	0.80
II	TT	25	0.08	0.09	0.14	0.07	***	0.10	0.14	0.18	0.36	0.70
	AR	25	0.07	0.08	0.13	0.06	0.02	***	0.10	0.12	0.32	0.71
	BB	25	0.07	0.07	0.14	0.04	0.03	0.02	***	0.09	0.29	0.64
	GA	25	0.09	0.09	0.15	0.05	0.04	0.02	0.01	***	0.27	0.66
III	BG	28	0.10	0.09	0.17	0.06	0.07	0.06	0.04	0.04	***	0.41
	AD	29	0.27	0.24	0.34	0.23	0.22	0.21	0.19	0.20	0.15	***
	Among geographical groups											
		I	II	III								
	I	***	0.24	0.58								
	II	0.07	***	0.40								
	III	0.15	0.10	***								

All values (both matrices) are significant after permutation test. *P* values were adjusted for a nominal 5% level for multiple comparisons

FR floristic region, *TB* Tarbagatay, *CR* Chernoff River, *KT* Krutoe truct, *LRB* Lepsy right bank, *TT* Tauturgen, *AR* Almaty reserve, *BB* Belbulak, *GA* Great Almaty gorge, *BG* Bozturgay gorge, *AD* Aksu Dzhabagly

of the cultivated apple (Cornille et al. 2014; Harris et al. 2002; Velasco et al. 2010). This species has contributed substantially to human's nutrition for thousands of years and contains a number of economically important traits urgently desired in many international apple breeding programs (Isutsa and Merwin 2000; Bassett et al. 2011; Wang et al. 2012; Janisiewicz et al. 2008). *M. sieversii* is not only native to Kazakhstan but also to Kyrgyzstan, Tajikistan, Uzbekistan, and China. The main part of Kazakhstan belongs to the Iran-Turan floristic region (Takhtadzhian 1986), which is characterized by its specific taxonomic, ecological and climatic conditions, geology, soil, and a multifaceted topography with large plains and hillocks, but also with high mountains of up to 7000 m covering about 7% of the total area. Each mountain ecosystem represents a separate floristic sub province which provides a suitable environment for many wild plants like *M. sieversii* (Flora of Kazakhstan 1956). Apple forests are distributed in the low mountain zones (steppes) at 800–1100 m, in middle mountains (forest steppes) at 1100–1500 m, and in forest-meadow-steppes at 1500–1800 m (Dzhangaliev 2003). Based on some changes in land use, *M. sieversii* is currently threatened with extinction in Kazakhstan (Eastwood et al. 2009). On this account, several plant explorations were made to collect plant propagation material (seeds, graft sticks) of *M. sieversii* in the center of its origin in Kazakhstan, Tajikistan, and Uzbekistan. Four of these explorations were sponsored by the National Plant Germplasm System of the USDA-Agricultural Research Service (Forsline et al. 2003). Between 1989 and 1996, American scientists collected *M. sieversii* at eight sites located

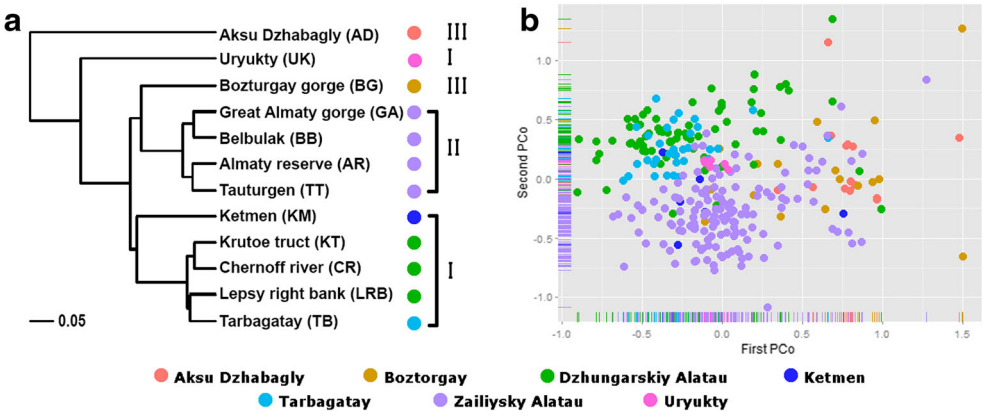

Fig. 3 UPGMA cluster analysis and principal coordinate analysis for the 311 *M. sieversii* genotypes. **a** Dendrogram obtained by UPGMA cluster analysis based on Nei's (1972) genetic distances, **b** with group membership defined accordingly to mountain ranges using principal coordinate analysis

in six different regions in Kazakhstan (Tarbagatay, Dzhungarskiy Alatay, Ketmen, Zailisky Alatau, Talas Alatau, and Karatau). Seeds of mother trees were brought to different international apple research institutions, where seedlings were planted and evaluated on their genetic variability among others (Richards et al. 2009b; Volk et al. 2005, 2013; Gross et al. 2013).

Since changes in land use are still proceeding, the present study aimed at investigating the current status of *M. sieversii* in Kazakhstan. Therefore, leaves of 311 individual plants of 12 *M. sieversii* wild apple populations natively growing in the eastern, south-eastern, and southern mountain regions of Kazakhstan were analyzed using a set of 16 SSR markers. All markers were polymorphic and informative with a mean PIC value of 0.75. Observed and expected heterozygosity were high suggesting a high level of diversity within the collected genotypes. According to Nei's algorithm, the present analysis revealed that the collection sites in the eastern, south-eastern, and southern mountainous regions of Kazakhstan are well differentiated with a genetic distance of 0.24 between sites I and II, 0.4 between II and III, and 0.59 between I and III. Despite the existing differences (e.g., half-sib families versus random seedlings in natural habitats, 8 versus 12 populations located at close but not the same coordinates (Table S4), a time interval of 25 years between both studies as well as 8 versus 16 microsatellite markers) between the experimental features published by Richards et al. (2009a) and here, both studies revealed high levels of Ho and He, different levels of within-site variation, as well as unique alleles, which is consistent with the previous analysis of *M. sieversii* populations.

The mean number of alleles per locus was high, but only slightly larger compared to Richards et al. (2009a), although a threefold number of trees was characterized in the present work versus 88 mother trees previously.

The samples of *M. sieversii* showed a high level of genetic diversity, and consistently low inbreeding coefficient. Genetic differentiation was measured using the F_{ST} index considering the genetic differentiation between the 12 populations. Low levels (F_{ST} <0.05) of differentiation were detected between the populations from eastern Kazakhstan (Tarbagatay (TB), Chernoff River (CR), Krutoe truct (KT), and Lepsy right bank (LRB)). Only between Krutoe truct and Ketmen (KM) was a moderate differentiation found. The observed within-population heterozygosity was between 0.62 and 0.71. A similar situation was observed for populations of the south-eastern region. The F_{ST} index was always below 0.05 for populations originating from the subprovince Zailiysky Alatau. This suggested a low differentiation between the populations, but relatively high within-population heterozygosity (0.67–0.75). A similar situation with a level of differentiation of 0.05 between sites was described by Richards et al. (2009a) while studying plant materials collected about 25 years ago. The level of differentiation between sites in the present study is similar to the findings of Richards et al. (2009a). Furthermore, the genetic distance was positively correlated with the geographic distance. Hence, both studies provide a good insight into the situation of *M. sieversii* at the time point of investigation. Richards et al. (2009a) argued that fragmentation and population isolation in the eastern and south-eastern areas of Kazakhstan may have occurred only recently from a large ancestral population as indicated by high levels of genetic diversity and the degree of genetic differentiation. Results of the present study are also in accordance with opinion of Cornille et al. (2012), that *M. sieversii* and *M. domestica* form distinctly separated groups with no significant differences in levels of genetic variation.

A significant number of hybrids between *M. sieversii* and *M. domestica* were detected for most of the populations ranging from 8 to 95%. This relatively high level of admixture suggests the occurrence of a frequent crop-to-wild gene flow. Significant levels of crop-to-wild gene flow between *M. domestica* and *M. sieversii* have previously been reported (Gross et al. 2012, Cornille et al. 2013). High rates of introgression suggest that hybrids are often viable and contribute to interspecific gene flow to a similar degree. The number of apple orchards surrounding wild populations and the total orchard area are positively related to recent introgression rates in wild apple populations (Cornille et al. 2015). The present situation of wild apple populations in Kazakhstan is the result of both current and past incidences. It is clear that the areas in the low mountain zones primarily in the piedmont-mountain border zone possessed the largest anthropogenic impact. The loss of biodiversity in these regions is due to the intensification of agriculture and urbanization. Moreover, in the 1960s and 1970s, the use of plots for private gardens in the mountainous areas was permitted (e.g., in Zailiysky Alatau). Plantings of cultivated apples in such private gardens increased the risk for crop-to-wild gene flow. For example, between 1932 and 1967, wild apples were often used as rootstocks directly at their place of growth and grafted with *M. domestica*. This additionally increased the risk of out crossing. Since these trees are still present everywhere in Kazakhstan, they continuously contribute to the crop-to-wild gene flow and increase the level of admixture in wild apple populations. Evidence for this is the 16 *M. sieversii* clones collected by Dzhangaliev collected in the wild, but which did not belong to *M. sieversii*. Whether these clones are hybrids or cultivated apple cultivars misclassified by Dzhangaliev is an entirely different debate. Nevertheless, based on their very low level of admixture with *M. sieversii*, it could be assumed they belong to *M. domestica*.

The high risk for crop-to-wild gene flow needs to be taken into account if future sample collection, evaluation, utilization, or conservation activities are planned (Gross et al. 2012). Especially pollen movement over long distances can detrimentally affect the genetic integrity of wild species.

Pollen movement in apple can lead to successful pollination events over distances of up to 10.7 km (Reim et al. 2015). These authors showed that a decrease in tree density resulted in an increase in pollen dispersal distances and the number of hybridization events. Conservation measures leading to an enhancement of the density of pure individuals will help to reduce the likelihood of hybridization. This can be achieved by repatriating plants from seed orchards, which were produced by controlled crossings between pure genotypes of the species of interest (Reim et al. 2015). On one hand, crop-to-wild gene flow is disadvantageous for the genetic integrity of wild species, but on the other hand, it could be a driving force for evolution. Hybridization between different species creates numerous new variations across genes and gene combinations. This increases the chance that individual genotypes with superior trait combinations will emerge, which will be better adaptable to the rapidly changing climatic conditions.

In conclusion, our results showed that nearly no admixture was found at Krutoe truct and Tauturgen. Both sites are located at altitudes of 1515 and 1585 m above sea level, respectively. We therefore recommend these sites for future in situ long-term preservation activities of mostly pure *M. sieversii* populations in Kazakhstan.

Acknowledgements The research was funded by the Grant on the subpriority "Fundamental Studies in the Area of Natural Sciences," Budget Program 101 of the Scientific Committee of the Ministry of Education and Science of the Republic of Kazakhstan (grant no. 1105/GF4). We gratefully acknowledge Dr. S.V. Chekalin for collecting *Malus sieversii* samples and providing information on the wild apple germplasm. We also acknowledge R. Gläß for her technical assistance and Dr. F.O. Emeriewen for improving the English language of the manuscript.

Data archiving statement All data are provided in Table S3.

References

Aubakirova K, Omasheva M, Ryabushkina N et al (2014) Evaluation of five protocols for DNA extraction from leaves of *Malus sieversii*, *Vitis vinifera* and *Armeniaca vulgaris*. Genet Mol Res 13(1):1278–1287

Bassett CL, Glenn DM, Forsline PL, Wisniewski ME, Farrell RE Jr (2011) Characterizing water use efficiency and water deficit responses in apple (*Malus* × *domestica* Borkh. and *Malus sieversii* Ledeb.) M. Roem. HortSci 46(8):1079–1084

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–331

Bykov BA (1961) In: Pavlov NV (ed) Flora of Kazakhstan [in Russian]. Kazakh Academy of Science, Alma-Ata, Volume 4, pp 402–405

Cornille A, Gladieux P, Smulders MJM, Roldan-Ruiz I, Laurens F, Le Cam B, Nersesyan A, Clavel J, Olonova M, Feugey L, Gabrielyan I, Zhang X-G, Tenaillon MI, Giraud T (2012) New insight into the history of domesticated apple: secondary contribution of the European wild apple to the genome of cultivated varieties. PLoS Genet 8(5):e1002703

Cornille A, Gladieux P, Giraud T (2013) Crop-to-wild gene flow and spatial genetic structure in the closest wild relatives of the cultivated apple. Evol Appl 6:737–748

Cornille A, Giraud T, Smulders MJ, Roldan-Ruiz I, Gladieux P (2014) The domestication and evolutionary ecology of apples. Trends Genet 30(2):57–65

Cornille A, Feurtey A, Gelin U, Ropars J, Misvanderbrugge K, Gladieux P, Giraud T (2015) Anthropogenic and natural drivers of gene flow in a temperate wild fruit tree: a basis for conservation and breeding programs in apples. Evol Appl 8:373–384

Cuizhi G, Spongberg SA (2003) Malus Miller. In: Wu Z, Raven PH (eds) Flora of China, volume 9. Science Press, Beijing, and Missouri Botanical Garden Press, St. Louis, pp 179–189

Dray S, Dufour AB (2007) The ade4 package: implementing the duality diagram for ecologists. J Stat Software 22(4):1–20

Dzhangaliev AD (1977) The wild apple tree of Kazakhstan. Nauka Publishing House of Kazakh SSR, Alma Ata

Dzhangaliev AD (2003) The wild apple tree of Kazakhstan. Hortic Rev 29:63–285

Earl DA, von Holdt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Res 4(2):359–361

Eastwood A, Lazkov G, Newton A (2009) The red list of trees of Central Asia. Fauna & Flora International, Cambridge, ISBN: 9781 903703 27 4

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

Flora of Kazakhstan (1956) Edited by Pavlov NV. KazSSR Academy of Science Publishing, Alma-Ata, V.1, pp 30–32

Forest encyclopedia (1986) Edited by Anuchin NP, Vorobiev GI. Soviet Encyclopedia, USSR, V.2., pp 180–182

Forsline PL, Aldwinckle HS, Dickson EE, Luby JJ, Hokanson SC (2003) Collection, maintenance, characterization, and utilization of wild apples of Central Asia. Hortic Rev 29:1–62

Goslee SC, Urban DL (2007) The ecodist package for dissimilarity-based analysis of ecological data. J Stat Software 22(7):1–19

Gross BL, Henk AD, Forsline PL, Richards CM, Volk GM (2012) Identification of interspecific hybrids among domesticated apple and its wild relatives. Tree Genet Genom 8(6):1223–1235

Gross BL, Volk GM, Richards CM, Henk AD, Forsline PL, Szewc-Mcfadden AK, Fazio G, Chao CT (2013) Diversity captured in the USDA-ARS National Plant Germplasm System apple core collection. J Am Soc Hortic Sci 138(5):375–381

Harris SA, Robinson JP, Juniper BE (2002) Genetic clues to the origin of the apple. Trends Genet 18(8):426–430

Hemmat M, Weeden NF, Brown SK (2003) Mapping and evaluation of *Malus* × *domestica* microsatellites in apple and pear. J Am Soc Hortic Sci 128(4):515–520

Hokanson SC, Szewc-McFadden AK, Lamboy WF, McFerson GR (1998) Microsatellite (SSR) markers reveal genetic identities, genetic diversity and relationships in a *Malus domestica* Borkh. Core subset collection. Theor Appl Genet 94:671–683

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. Mol Ecol Res 9(5):1322–1332

Ignatov A, Bodishevskaya A (2011) Malus. In: Kole C (ed) Wild crop relatives: genomic and breeding resources: temperate fruits. Springer-Verlag, Berlin Heidelberg, pp 45–64

Isutsa DK, Merwin IA (2000) Malus germplasm varies in resistance or tolerance to apple replant disease in a mixture of New York orchard soils. HortSci 35(2):262–268

Ivashenko AA (2005) The treasures of flora in Kazakhstan. Through the pages of the red Book. Almaty Kitap, Almaty, p 119

Jakobsson M, Noah A (2007) Rosenberg CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23(14):1801–1806

Janisiewicz WJ, Saftner RA, Conway WS, Forsline PL (2008) Preliminary evaluation of apple germplasm from Kazakhstan for resistance to postharvest blue mold in fruit caused by Penicillium expansum. HortSci 2(43):420–426

Jombart T (2008) Adegenet: an R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405

Kalinowski ST (2005) HP-Rare: a computer program for performing rarefaction on measures of allelic diversity. Mol Ecol Notes 5: 187–189

Kamvar ZN, Brooks JC, Grünwald NJ (2015) Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. Front Genet 6:208

Liebhard R, Gianfranceschi L, Koller B, Ryder CD, Tarchini R, Van de Weg WE, Gessler C (2002) Development and characterization of 140 new microsatellites in apple (*Malus* x *domestica* Borkh.) Mol Breed 10:217–241

Matschiner M, Salzburger W (2009) TANDEM: integrating automated allele binning into genetics and genomics workflows. Bioinformatics 25(15):1982–1983

Missiaggia A, Grattapaglia D (2006) Plant microsatellite genotyping with 4-color fluorescent detection using multiple-tailed primers. Genet Mol Res 5:72–78

Nei M (1972) Genetic distance between populations. Am Nat 106(949): 283–292

Paradis E, Claude J, Strimmer K (2004) APE: analyses of phylogenetic and evolution in R language. Bioinformatics 20:289–290

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in excel: population genetic software for teaching and research update. Bioinformatics 19:2537–2539

Pons L (2006) Remarkable Kazak apples. AgResearch Magazine 54(1): 4–6

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

R Development Core Team (2008) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org

Reim S, Proft A, Heinz S, Lochschmidt F, Höfer M, Tröber U, Wolf H (2015) Pollen movement in a *Malus sylvestris* population and conclusions for conservation measures. Plant Genet Res 1–9. doi: 10.1017/S1479262115000301

Richards CM, Volk GM, Reilley AA, Henk AD, Lockwood DR, Reevers PA, Forsline PL (2009a) Genetic diversity and population structure in *Malus sieversii*, a wild progenitor species of domesticated apple. Tree Genet Genomes 5:339–347

Richards CM, Volk GM, Reeves PA, Reilley AA, Henk AD, Forsline PL, Aldwinckle HS (2009b) Selection of stratified core sets representing wild apple (*Malus sieversii*). J Am Soc Hortic Sci 134(2):228–235

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

Severtsov NA (1873) Travel on Turkestan region and research of mountain country Tien Shan. Accomplished on a Commission of Russian Geographic Society, St. Petersburg, pp 82–83

Takhtadzhian AL (1986) Floristic regions of the world. University of California Press, Berkley, ISBN-10: 0520040279, ISBN-13: 978–0520040274

Turehanova RM (2012) Sievers apple tree in Kazakhstan, research and selection of forms. Terrain 12(1):44–61

Vavilov NI (1931) The wild relatives of fruit trees of the Asian part of the USSR and Caucasus and the problem of the origin of fruit trees. Trans Appl Bot Gene Breed 26(3):132–134

Velasco R, Zharkikh A, Affourtit J et al (2010) The genome of the domesticated apple (*Malus* × *domestica* Borkh.) Nature Genet 42:833–839

Volk GM, Reilley AD, Henk PL, Forsline HS, Aldwinckle, Richards CM (2005) Ex situ conservation of vegetatively-propagated species: development of a seed-based core collection for *Malus sieversii*. J Am Soc Hortic Sci 130:203–210

Volk GM, Richards CM, Henk AD, Reilley AD (2009) Novel diversity identified in a wild apple population from the Kyrgyz Republic. HortSci 44(2):516–518

Volk GM, Henk AD, Richards CM, Forsline PL, Chao CT (2013) *Malus sieversii*: a diverse Central Asian apple species in the USDA-ARS national plant germplasm system. HortSci 48:1440–1444

Wang A, Aldwinckle H, Forsline P, Main D, Fazio G, Brown S, Xu K (2012) EST contig-based SSR linkage maps for *Malus* ×*domestica* cv Royal Gala and an apple scab resistant accession of *M. sieversii*, the progenitor species of domestic apple. Mol Breed 29:379–397

Wickham H (2009) ggplot2: elegant graphics for data analysis. Springer-Verlag, New York

Zhang C, Chen X, He T, Liu X, Feng T, Yuan Z (2007) Genetic structure of *Malus sieversii* population from Xinjiang, China, revealed by SSR markers. J Genet Genomics 34:947–955