Tree Genetics & Genomes (2009) 5:339–347
DOI 10.1007/s11295-008-0190-9

ORIGINAL PAPER

Genetic diversity and population structure in *Malus sieversii*, a wild progenitor species of domesticated apple

Christopher M. Richards · Gayle M. Volk · Ann A. Reilley · Adam D. Henk · Dale R. Lockwood · Patrick A. Reeves · Philip L. Forsline

Received: 29 November 2007 / Revised: 3 September 2008 / Accepted: 4 October 2008 / Published online: 9 December 2008
© US Government 2008

Abstract *Malus sieversii* (Lebed.) M. Roem. is a wild progenitor species of the domesticated apple. It is found across a mountainous region of central Asia and has been the focus of several collection expeditions by the USDA-ARS-National Plant Germplasm System. This study used microsatellite variation at seven loci to estimate diversity and differentiation within *M. sieversii* using several complimentary approaches. Multilocus genotypes were amplified from 949 individuals representing seedling trees from 88 half-sib families from eight *M. sieversii* populations collected in Kazakhstan. Apportioning of genetic variation was estimated at both the family and site level. Analyses using a hierarchical model to estimate F_{st} showed that differentiation among individual families is more than three times greater than differentiation among sites. In addition, average gene diversity and allelic richness varied significantly among sites. A rendering of a genetic network among all sites showed that differentiation is largely congruent with geographical location. In addition, nonhierarchical Bayesian assignment methods were used to infer genetic clusters across the collection area. We detected four genetic clusters in the data set. The quality of these assignments was evaluated over multiple Markov Chain Monte Carlo runs using both posterior likelihood and stability of the assignments. The spatial pattern of genetic assignments among the eight collection sites shows two broadly distributed and two narrowly distributed clusters. These data indicate that the southwestern collection sites are more admixed and more diverse than the northern sites.

Keywords Genetic diversity · *Malus* · Genetic resources

Communicated By: A. Dandekar

C. M. Richards (✉) · G. M. Volk · A. A. Reilley · A. D. Henk · P. A. Reeves
National Center for Genetic Resources Preservation,
United States Department of Agriculture,
Fort Collins, CO 80521, USA
e-mail: christopher.richards@ars.usda.gov

D. R. Lockwood
Department of Biology,
Colorado State University,
Fort Collins, CO 80523, USA

P. L. Forsline
Plant Genetic Resources Unit,
United States Department of Agriculture,
Geneva, NY 14456-0462, USA

Introduction

Many domestic apple genotypes trace their ancestry to a small number of progenitor lineages and, therefore, share a high degree of genetic identity (Noiton and Alspach 1996). The narrowness of this genetic base has been further eroded as the number of genotypes used in commercial production has decreased over the last century (Hokanson et al. 2001). In contrast, wild species within this genus are known to contain substantial genetic diversity. In particular, *Malus sieversii* (Lebed.) M. Roem. is recognized as a major progenitor species to the cultivated apple, *M.* × *domestica* Borkh (Forte et al. 2002; Watkins 1995; Zhou and Li 2000 but also see Coart et al 2006; Harris et al. 2002). Wild populations of *M. sieversii* exist as forested stands

throughout a remote mountainous region of the Tien Shan Mountains in Central Asia (Morgan and Richards 1993). The diversity of individual phenotypes from the native landscape is remarkable and was the basis for Vavilov's conclusion that Kazakhstan and Central Asia are the center of origin for apples (Vavilov 1992). Fruits of *M. sieversii* are the largest of any wild species of *Malus* and range from large edible fruits to fruits that are small and astringent (Luby et al. 2001). Most importantly, *M. sieversii* populations display diverse phenotypic characters that represent a critical genetic resource for disease resistance, fruit quality, and tree physiology in cultivated apple (Forsline and Aldwinkle 2004, Luby et al. 2002; Volk et al. 2005).

Four plant explorations were sponsored by the USDA-Agricultural Research Service's National Plant Germplasm System to 12 regions of Kazakhstan, Tajikistan, and Uzbekistan between 1989 and 1996 (Forsline et al. 2003). As an insect-pollinated, outcrossing, perennial species, *M. sieversii* is expected to show modest to low differentiation and genetic structuring among populations (Hamrick and Godt 1996; Morgan and Richards 1993). However, given the size and edibility of the fruits, the genetic structure of this species may be under considerable human influence. The native range of *M. sieversii* lies within significant overland trade routes between Asia and Europe, and dispersal may, in part, be influenced by this activity (Harris et al. 2002; Dzhangaliev 2003; Luby et al. 2001).

Because of the uncertainty of estimating dispersal and genetic structure in a species that may have been influenced by anthropogenic activities, we used several complimentary approaches to estimate population structure and to examine broad patterns of genetic admixture and differentiation in this species. Assessments of genetic diversity often make use of information about the current geographic relationships of individuals. These estimators are excellent for investigating patterns within and among predefined sampling hierarchies. However, recent research using large genotypic data sets has shown that genetic discontinuities may not adequately fit a priori assumptions of defined spatial boundaries (Nordborg et al. 2005; Rosenberg et al. 2002). One way around this limitation is to partition genotypes into genetic clusters using linkage disequilibrium among loci as an assignment criterion without using any prior spatial information (Pritchard et al. 2000, Waples and Gaggiotti 2006). The advantage of this approach is that the clusters of genotypes drawn from different locations allows for the assessment of admixture and may provide a flexible way of identifying patterns of differentiation at regional scales.

Here, we examine the genetic diversity and differentiation of 949 individuals collected from eight locations within Kazakhstan genotyped for seven microsatellite loci. Our goal is to assess genetic diversity and to help support conservation, management, and utilization of this important wild germplasm for apple improvement.

Materials and methods

Plant materials

M. sieversii plant materials were collected from 12 sites within the Central Asian countries of Kazakhstan, Uzbekistan, and Tajikistan during exploration visits in 1989, 1993, 1995, and 1996. In Kazakhstan, seeds were collected from as far west as the Karatau mountain range (sites 6, 11, and 12) and in the eastern part of Kazakhstan from the Tarbagatai mountain range (site 9), the Djungarsky mountain range (sites 4 and 5), near Ketmen (site 10), and near Almaty (site 3). The terrain and elevation of the sites varied (Table 1), and detailed site descriptions are presented by Dzhangaliev (2003) and Forsline et al. (2003). Maternal trees were randomly selected, and in most cases, seeds collected from a single maternal tree were considered a unique seed lot.

Seeds were brought through quarantine to the USDA-Agricultural Research Service (ARS) Plant Genetic Resources Unit (PGRU) located in Geneva, NY. Between one and 30 seeds were planted to represent selected seed lots from the Kazakhstan collection sites. Seedlings were grown under greenhouse conditions and then transplanted to field orchards. Each seedling was given a GMAL (Geneva-*Malus*) identification number corresponding to their family seed lot and a unique individual suffix. Eight sites had adequate quantities of seeds to be included in analyses within this report. Leaf samples were collected from each *M. sieversii* tree in 2002 and sent to the USDA-ARS National Center for Genetic Resources Preservation (NCGRP) in Ft. Collins, CO and kept at −80°C until DNA extractions were performed.

Molecular analysis

Genomic DNA from leaf tissue of 949 individual *M. sieversii* trees was extracted using DNeasy 96 plant kits (Qiagen, Valencia, Calif.). Two samples from each individual were extracted and polymerase chain reactions (PCR) amplified independently to confirm results. *Malus* simple sequence repeats (SSR) were amplified using unlinked primers (GD12, GD15, GD96, GD100, GD142, GD147, GD162) as described by Hokanson et al. (1998) and Hemmat et al. (2003). Forward primers, labeled with either IRD 700 or IRD 800, were obtained from MWG-Biotech (High Point NC). Unlabeled reverse primers were purchased from IDT (Coralville, IA, USA). All PCR were carried out in 15 μl total volume using previously published

Table 1 Physical descriptions of the eight Kazakhstan collection sites included in microsatellite analyses

Collection site	Region	Collection date	Habitat description	Latitude	Longitude	Elevation (m)	Precipitation (mm)
3	Zailisky	1993, 1995, 1996	Humid-temperate, mixed forest	43°06′–43° 21′ N	77°23′–77° 54′ E	1,170–1,690	700
4	Djungarsky	1993, 1995, 1996	Humid-temperate, mixed forest	45°31′–45° 41′ N	80°43′–80° 55′ E	1,190–1,360	800
5	Djungarsky	1993, 1995, 1996	Humid-temperate, mixed forest	45°24′ N	80°24′–80° 25′ E	1,170–1,760	850
6	Karatau	1993, 1995	Stream habitat and xeric, mixed scrub forest	42°52′ N	69°53′ E	600–910	250
9	Tarbagatai	1995, 1996	Dry continental forest with −40°C common	47°14′–47° 16′ N	81°34′–81° 35′ E	890–1,120	450
10	Ketmen	1996	Semidry, temperate, mixed forest	43°18′ N	79°31′ E	1,600–1,700	650
11	Karatau	1996	Xeric area and stream habitat	42°40′N	70°16′ E	780–1,230	250
12	Talasky	1996	Dry north slope of canyon	42°19′ N	70°22′ E	1,000–1,025	320

methods (Volk et al. 2005). PCR products were visualized and scored on a Licor 4200 DNA Sequencer (Licor, Lincoln, NE, USA). Digital images were collected from the sequencer using Licor Saga™ Generation 2 software and were manually interpreted and scored using the Saga™ software. Alleles from replicate samples were examined at each locus, and when alleles for replicates were not identical, data for that locus were entered as "missing" in subsequent analysis. Individuals were included in the analyses when they had missing data for no more than one marker.

Data analysis

Of the 949 genotypes in the data set, 922 were from families that contain five or more individuals. Statistics were performed using two one-way analyses or two-level analyses where families were nested within sites. In our analysis, sites and families could not be nested within genetic cluster. Descriptive statistics, including expected and observed heterozygosities, differentiation between groups (overall and pairwise F_{st}), and diversity within groups as measured by the absolute number of polymorphic alleles and a sample adjusted metric of allelic richness useful for group comparisons (El Mousadik and Petit 1996), were estimated from genotypic data using the software package GDA (Lewis and Zaykin 2001) and *Fstat* (Goudet 1995).

Nonhierarchical genotypic clustering was performed using the genotypes obtained for all 949 samples in a manner independent of site or family structure. This clustering approach uses genetic linkage disequilibrium (LD) among alleles at each of the marker loci and partitions genotypes to minimize LD within the clusters. Inferring the number of clusters (denoted k) in a data set is often challenging and is affected by the number of loci used, the variation at each of these loci, the scale and magnitude of gene flow, and the number of samples in the data set (Evanno et al. 2005). We estimated the number of clusters and the assignment of individuals into clusters using three complimentary methods.

The number of clusters was first estimated by determining the change in the posterior probability of the assignments over different values of k. We chose an ancestry model that assumes admixture and correlated allele frequencies between populations where the posterior probability is evaluated by fractionally assigning individual genotypes to multiple clusters using a membership coefficient. Posterior probabilities were computed in *Structure* (Pritchard et al. 2000) using an MCMC method; over 100,000 iterations of each chain following a 500,000 iteration burn-in period. Each Markov chain was run 100 times for k=1–12. We implemented the Δk method described by Evanno et al. (2005) to examine the rate of change in successive posterior probabilities over a range of k values. This approach takes into consideration both the gain in posterior probabilities and the variance among independent runs at these values. A genotype was assigned to the cluster for which it had the highest membership coefficient. The modal cluster assignment across all replicate runs was used to determine the final placement of a genotype in a cluster.

Secondly, the number of clusters and the assignment of genotypes to those clusters were estimated simultaneously by treating k as a random variable that follows a Dirichlet process prior distribution using the software *Structurama* (Huelsenbeck and Andolfatto 2007). Each analysis consisted of a single Markov chain that was allowed to run for 100,000 iterations. A partition was sampled from the Markov chain every 100 iterations following a burn-in

period of 50,000 iterations. The mean partition, which minimizes the squared distance to the sampled partitions, was calculated in order to make assignments.

Finally, a stability coefficient (S_N) was computed for each of the two methods across 100 independent runs. Matrix M represents the pairwise correlations of individuals among N runs. The value at the ith row and jth column is the number of times that individuals i and j have the highest membership coefficient (Q value) for the same cluster, i.e., they are preferentially assigned to the same cluster. Each value in M is divided by the number of runs, N, to generate a frequency. M is the upper triangular portion of this pairwise matrix to avoid both duplication and the trivial case of M_{ii}. The stability coefficient is thus given by,

$$S_N = \frac{\sum_{i=l}^{n}\sum_{j>i}^{n} M_{ij}}{q}$$

where n is the number of individuals sampled, and q is the number of nonzero entries in M. The k value that maximizes S_N was preferred. This stability coefficient and the similarity coefficient of Rosenberg et al. (2002) report of the same basic correlation, but the stability coefficient reported here has the added advantage of being computable for large k values and large numbers of runs (Lockwood unpublished results).

Networks are useful in displaying relationships among groups that have reticulate histories (gene flow) and do not strictly meet the assumptions of hierarchical clustering methods. To visualize the genetic relationships among sampling sites in Kazakhstan, we used the methods developed by Dyer and Nason (2004) that use a graph theoretical approach that is free of hierarchical restrictions and can accommodate multilocus genotype data. In this rendering, genotypes grouped by site of collection are displayed in a network topology rather than a bifurcating dendrogram. In the graph, sampling strata are represented as nodes whose diameter is proportional to the within strata multilocus genetic variance (Dyer and Nason 2004). Edge lengths represent the among-population component of genetic variation and are proportional to the degree of differentiation among the nodes. Node relationships that are significantly structured are not displayed as a complete graph (where all pairwise $n(n-1)/2$ connections are drawn as edges). Instead, only the smallest subset of edges necessary to maintain the overall genetic covariance among the populations is drawn. In graph theory, nodes that are connected to all other $n-1$ nodes in the graph have the highest edge degree. Genetic structuring is tested with a chi-squared significance test to determine if there are deficiencies of edges between nodes relative to the saturated graph.

Results

The seven microsatellites used in this study were informative, amplifying a total of 103 alleles (Table 2). The number of polymorphic alleles among all 949 individuals ranged from 3 (GD15) to 29 (GD162). No duplicate genotypes were identified. Both observed and expected heterozygosity values across each locus were high suggesting that the collection of genotypes was highly diverse (Table 2).

Variation among collection sites

In our sampling design, a significant part of the genetic structure of *M. sieversii* is explained by variation among individual families that corresponds to seedlings taken from individual maternal trees. The average estimates of F_{st} at the family and site levels were 0.19 and 0.05, respectively, when two-level models are run with family nested within site (Table 3). Significant structure at the family level is not surprising and may arise from long generation times and low infrequent levels of population recruitment. In most families, we did not have enough individuals to accurately estimate the outcrossing rates among maternal trees, but overall inbreeding coefficients among families showed a significant excess of heterozygosity (F_{is}=−0.134), which is consistent with outcrossing in these family groups. Estimates of excess heterozygosity were significant using a Wilcoxon signed rank test, but these results need to be treated cautiously, since sample sizes for families ranged from five to 29 individuals with an average of ten. In contrast, F_{is} at the genetic cluster level (F_{is}=0.022) was not statistically different than zero (Table 3).

The average F_{st} among sites was 0.05, suggesting that there are small, but significant differences among sampling sites (Table 3). Pairwise F_{st} estimates were all significantly nonzero using a permutation test (Table 4). Site-specific diversity measurements revealed that allelic richness varied

Table 2 Descriptive statistics by locus

Marker	Polymorphic Alleles (no.)	He	Ho	*n*
GD12	16	0.794	0.811	931
GD15	3	0.497	0.449	943
GD96	20	0.842	0.817	941
GD100	9	0.590	0.377	859
GD142	14	0.862	0.823	947
GD147	12	0.849	0.808	946
GD162	29	0.809	0.768	946
Total	103			
Average/locus	14.7	0.749	0.693	

Expected heterozygosity (He), observed heterozygosity (Ho), and number of scored genotyped (n) are provided for each microsatellite marker

Table 3 Average measures of genetic differentiation including inbreeding coefficient of the subpopulation (F_{is}), inbreeding in the total sample (F_{it}), and genetic differentiation (F_{st}) among sites and families within sites and genetic clusters identified through Bayesian assignment methods

Locus	F_{is}	F_{it}	F_{st} families	F_{st} sites	F_{st} cluster
Among families (within sites)					
GD12	−0.223	−0.015	0.170	0.038	
GD15	−0.242	0.116	0.288	0.188	
GD96	−0.154	0.039	0.167	0.030	
GD100	0.147	0.369	0.261	0.023	
GD142	−0.148	0.053	0.176	0.041	
GD147	−0.130	0.062	0.170	0.026	
GD162	−0.139	0.058	0.173	0.052	
Overall	−0.134	0.084	0.192	0.052	
Upper bound	−0.051	0.179	0.232	0.087	
Lower bound	−0.187	0.029	0.171	0.030	
Among genetic clusters					
GD12	−0.056	0.003			0.056
GD15	−0.069	0.192			0.244
GD96	−0.011	0.057			0.068
GD100	0.298	0.407			0.156
GD142	0.022	0.061			0.039
GD147	0.009	0.077			0.069
GD162	0.009	0.079			0.070
Overall	0.022	0.111			0.090
Upper bound	0.105	0.212			0.144
Lower bound	−0.028	0.049			0.056

considerably across sites (Table 5). The most diverse sites were 5, 11, and 12. In addition, there were a number of alleles that were only found in certain sites. Most of these private alleles were found at low frequency (below 0.05), but some, such as allele 140 at locus GD12 and allele 153 at locus GD147, were common (around 0.5) in site three (data not shown).

A population graph shows that each site has multiple connections to other sites and a number of triangular loops that are consistent with reticulate gene flow among the collection sites and low level of average F_{st} values (Fig. 1). While the topology shows no marked subdivision such as distinct subgraphs connected by a single node, there is broad correspondence to the geographical location of each site. Nodes 4, 5, and 9, corresponding to sites located in the northeastern regions of Kazakhstan, are located along the right-hand margin of the graph. Nodes 6, 11, and 12, representing the western collection sites, are along the left-

Table 4 Pairwise differentiation (F_{st}) among collection sites and genetic clusters

	2	3	4	5	6	9	10	11	12
Among collection sites									
3			0.0821	0.0589	0.0886	0.0649	0.0679	0.1284	0.124
4				0.0412	0.1162	0.0391	0.0912	0.1384	0.1519
5					0.056	0.0199	0.0361	0.0685	0.0734
6						0.0726	0.0475	0.0374	0.0658
9							0.0434	0.0833	0.0885
10								0.0642	0.0626
11									0.0457
Among genetic clusters									
1	0.0562	0.1086	0.2098						
2		0.0948	0.1461						
3			0.146						
4									

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

Table 5 Allelic richness per locus and site and per locus and cluster

Locus	1	2	3	4	5	6	9	10	11	12
Site[a]										
GD12			6.92	6.73	6.83	7.51	6.05	6.80	7.22	7.56
GD15			2.00	2.00	2.29	2.00	2.00	2.00	2.59	2.00
GD96			8.92	5.24	8.55	9.85	6.48	8.14	10.41	9.73
GD100			4.00	4.33	4.98	4.15	3.38	3.70	3.93	4.99
GD142			6.87	7.84	9.99	6.73	9.02	8.75	8.11	9.58
GD147			8.96	6.78	7.49	7.64	8.04	7.63	8.83	8.16
GD162			10.92	8.27	11.80	10.34	9.71	8.85	10.87	9.08
Mean/site			6.94	5.88	7.42	6.89	6.38	6.55	7.42	7.30
Private alleles			2	0	0	0	4	0	5	4
N			24	93	142	173	252	59	117	89
Cluster[b]										
GD12	7.65	8.52	7.65	9.00						
GD15	2.25	2.12	2.00	2.00						
GD96	8.29	10.56	11.10	12.72						
GD100	4.70	4.94	4.07	3.83						
GD142	10.55	7.06	9.96	10.82						
GD147	8.43	7.51	8.62	8.83						
GD162	12.98	11.31	12.40	14.31						
Mean/Cluster	6.98	6.79	7.23	7.87						
Private Alleles	9	6	5	4						
N	561	258	98	32						

[a] Data per locus is based on a minimum sample size of 23 diploid individuals
[b] Data per locus is based on a minimum sample size of 31 diploid individuals

hand margin of the graph. Finally, nodes representing sites 3 and 10 both show strong linkages with both northeastern and western collection site nodes. Variation within the sites (Table 5) is correlated with node diameter but is not consistently correlated with sample size. For instance, node 3 is the smallest site collection (32 individuals) and yet shows the highest edge degree, a central location in the network topology, as well as greater gene diversity (node diameter) and allelic richness (Table 5) than site 6 (173 individuals). The two smallest network nodes correspond to sites 4 and 9, yet the number of samples from these sites are intermediate and large (93 and 252, respectively).

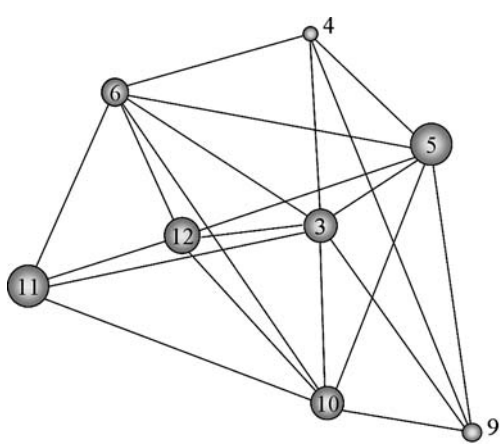

Fig. 1 Population network of collection sites in Kazakhstan. Nodes represent each of the eight collection sites. Node diameter is proportional to intrasite genetic variation. Length of edges connecting nodes is proportional to genetic differentiation among the connected sites (see text for details). Both node sizes and edge lengths are rendered in a three-dimensional space

Variation by cluster

Individuals were clustered according to their genotypes, without consideration of collection or maternal origin. Models that maximized the highest posterior likelihood and minimized variance among *Structure* runs had k=4 clusters. Using the Δk method, we found the largest change in posterior probability to be between k=3 and k=4. The *Stucturama* analysis also converged on four well-supported groups. The stability coefficient at this k=4 value (0.899) was the highest across a range of values k=3 to 15. Average genetic differentiation among all four groups was significant using a permutation test (Table 4), but this may be due to the fact that the size distribution among each cluster was highly skewed—from cluster 1 with 561 individuals to cluster 4 with 32 individuals (Table 5). The largest two clusters (1 and 2) were less differentiated from each other than to the other two smaller clusters. As in the site analysis, allelic richness within each cluster did not correlate with the cluster size (Table 5). The distribution of cluster assignments for

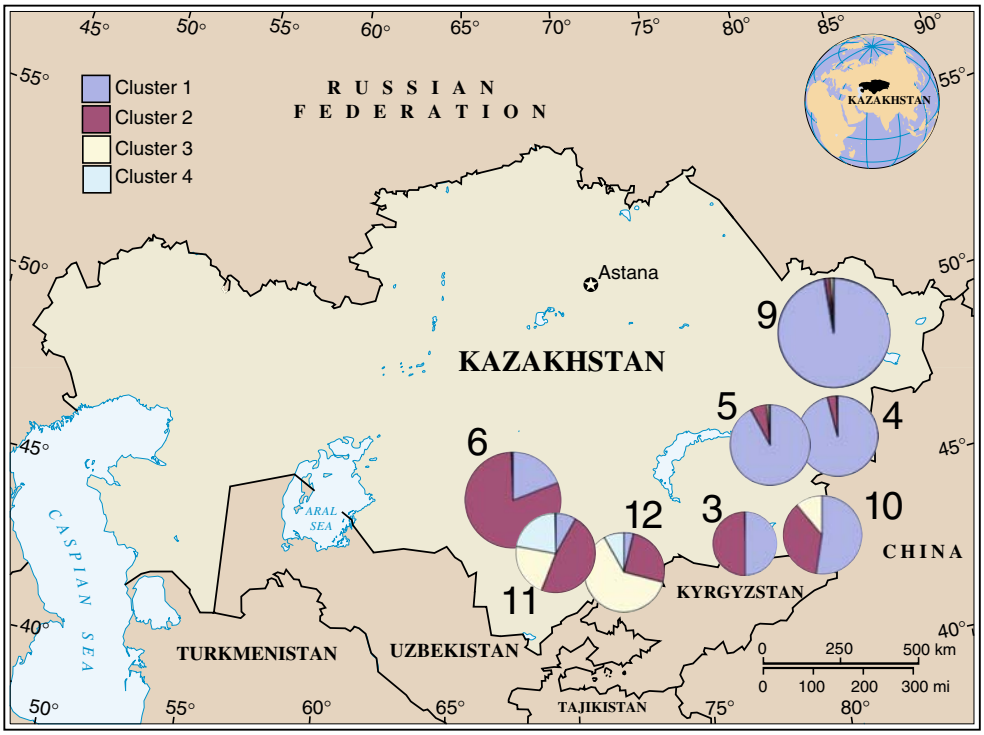

Fig. 2 Geographic pattern of genetic clusters in Kazakhstan. *Pie charts* located at each of the eight collections sites reflect the proportion of individuals assigned to each of the four genetic clusters identified through Bayesian analysis. Node sizes correlate roughly with sample size at each site

individuals from each site was nonrandom with respect to geographical location (Fig. 2). Individuals composing cluster 1 were largely drawn from locations in the north and central collection regions such as sites 3, 4, 5, 9, and 10. Cluster 2 showed a broad distribution throughout the central and southern regions with the highest concentration in site 6 and 11. Both clusters 3 and 4 are largely restricted to the southern collection regions in sites 11 and 12. The genetic diversity within clusters 3 and 4 was higher than clusters 1 and 2 (Table 5).

Discussion

The sampled *M. sieversii* populations in Kazakhstan are clearly not panmictic which contrasts previous estimates based on fewer samples for this region (Lamboy 1996). Analyses revealed that the collection sites are significantly differentiated and have different levels of within-site variation, as well as unique alleles. The regional pattern of diversity revealed by genetic cluster assignments depicts a more subtle regional structure that may be particularly helpful in helping to support genebank management of the existing collection and guiding efforts to collect additional samples. These inferences could not have been estimated from *a priori* population designations, and they point out a broad pattern of biogeographical patterning across this region. Admixture increased from the northwest collection sites to the southwestern collection sites. These results are supported by previous reports of phenotypic variation such as disease resistance among sites (Forsline and Aldwinckle 2004; Forsline et al. 2003). Additional sampling in the neighboring areas of this species range may help to identify the ecological and demographic influences that shape this population structure.

The use of molecular marker data has been extremely useful for organizing and managing collections of genetic resources (Bretting and Widrlechner 1995; Van Treuren et al. 2001). In particular, genetic analyses have been performed with microsatellite markers in cultivated and wild *Malus* collections for fingerprinting (Galli et al. 2005; Guilford et al. 1997; Hokanson et al. 1998; Oraguzie et al. 2005), genomic mapping (Hemmat et al. 2003; Liebhard et al. 2002; Liebhard et al. 2003), marker-assisted breeding (Gianfranceschi et al. 1999; Gygax et al. 2004), and assessments of diversity within domesticated and wild collections (Coart et al. 2003; Gaurino et al. 2006; Hokanson et al. 2001; Kitahara et al. 2005; Lamboy et al. 1996; Pereira-Lorenzo et al. 2003; Ramos-Cabrer et al. 2007; Volk et al. 2005; Volk et al. 2008; Yan et al. 2008; Zhang et al. 2007).

In this study, the average number of alleles per locus (14.7) was similar to that found by Hokanson et al. (1998) when eight (seven of which are used in this study) microsatellite markers were used to compare 66 *Malus* × *domestica* cultivars. The mean expected heterozygosity (0.749) was higher than the 0.693 reported by Hokanson et al. (1998). In both the current research and the Hokanson

et al. (1998) publication, GD15 had the fewest number of alleles (3 and 2, respectively). Not surprisingly, higher levels of gene diversity were found by Hokanson et al. (2001) when 142 accessions representing a broad spectrum of 23 species and hybrids were analyzed using the same microsatellite markers.

Currently, 1,198 individuals from the *M. sieversii* seed collections are established in a permanent orchard managed by the ARS-Plant Genetic Resources Unit, Geneva, NY. These accessions have been evaluated for a number of valuable horticultural traits including resistance to fire blight (*Erwinia amylovora* Burrill), apple scab resistance (*Venturia inaequalis* Cooke), fruit quality, and drought tolerance. A collection of mature trees in the field is useful for quantitative trait assessment, but comes with inherent risk of mortality. For instance 17% of the initial plantings have died, and an additional 15% appear to be weak. At an applied level, understanding the pattern of genetic variation for both phenotypic and molecular traits will help develop seed-based core collections (Volk et al. 2005) from these orchards for secure long term of this important genetic resource.

Acknowledgements We thank Dr. Angela Baldo and Dr. Gennaro Fazio of the ARS Plant Genetic Resources Unit in Geneva, NY for helpful comments on earlier drafts of this work. This project was partially supported by the National Research Initiative of the USDA Cooperative State Research, Education and Extension Service, grant number #2005-00751. Dale Lockwood was supported by a USDA-ARS Specific Cooperative Agreement with Colorado State University's Program for Interdisciplinary Mathematics, Ecology and Statistics. Any mention of trade names of commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

References

Bretting PK, Widrlechner MP (1995) Genetic markers and plant genetic resources. Plant Breed Rev 13:11–86

Coart E, Vekemans Z, Smulders M, Wagner I, van Huylenbroeck J, van Bockstaele E, Roldan-Ruiz I (2003) Genetic variation in the endangered wild apple (*Malus sylvestris* (L.) Mill.) in Belgium as revealed by amplified fragment length polymorphism and microsatellite markers. Mol Ecol 12:845–857

Coart E, van Glabeke S, De Loose M, Larsen AS, Roldan-Ruiz I (2006) Chloroplast diversity in the genus *Malus*: new insights into the relationship between the European wild apple (*Malus sylvestris* (L.) Mill.) and the domesticated apple (*Malus domestica* Borkh.). Mol Ecol 15:2171–2182

Dyer RJ, Nason JD (2004) Population graphs: the graph-theoretic shape of genetic structure. Mol Ecol 13:1713–1728

Dzhangaliev AD (2003) The wild apple tree of Kazakhstan. Hort Rev 29:63–303

El Mousadik A, Petit RH (1996) High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic of Morocco. Theor Appl Genet 92:832–839

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

Forsline PL, Aldwinckle HS (2004) Evaluation of *Malus sieversii* seedling populations for disease resistance and horticultural traits. Acta Hort 663:529–534

Forsline PL, Aldwinckle HS, Dickson EE, Hokanson SC (2003) Collection, maintenance, characterization, and utilization of wild apples from central Asia. Hort Rev 29:1–61

Forte AV, Ignatov AN, Ponomarenko VV, Dorokhov DB, Savelyev NI (2002) Phylogeny of the *Malus* (apple tree) species, inferred from the morphological traits and molecular DNA analysis. Russian J Genet 38:1150–1160

Galli Z, Halasz G, Kiss E, Heszky L, Dobranszki J (2005) Molecular identification of commercial apple cultivars with microsatellite markers. HortScience 40:1974–1977

Gaurino C, Santoro S, De Simone L, Lain O, Cipriani G, Testolin R (2006) Genetic diversity in a collection of ancient cultivars of apple (*Malus* X *domestica* Borkh.) as revealed by SSR-based fingerprinting. J Hort Sci Biotechnol 81:39–44

Gianfranceschi L, Seghas N, Kellerhals M, Gessler C (1999) Molecular markers applied to apple breeding: Analysis of oligogenic and single gene resistances. Acta Hort 484:417–428

Goudet J (1995) FSTAT, a program for IBM PC compatibles to calculate Weir and Cockerham's (1984) estimators of F-statistics. J Hered 86:485–486

Guilford P, Prakash S, Zhu JM, Rikkerink E, Gardiner S, Bassett H, Forster R (1997) Microsatellites in *Malus* X *domestica* (apple): abundance, polymorphism and cultivar identification. Theor Appl Genet 94:249–254

Gygax M, Gianfranceschi L, Liebhard R, Kellerhals M, Gessler C, Patocchi A (2004) Molecular markers linked to the apple scab resistance gene *Vbj* derived from *Malus baccata jackii*. Theor Appl Genet 109:1702–1709

Hamrick JL, Godt MJW (1996) Effects of life history traits on genetic diversity in plant species. Philosophical Trans Royal Soc London Series B Biol Sci 351:1291–1298

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

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

Hokanson SC, Lamboy WF, Szewc-McFadden AK, McFerson JR (2001) Microsatellite (SSR) variation in a collection of *Malus* (apple) species and hybrids. Euphytica 118:281–294

Hokanson SC, Szewc-McFadden AK, Lamboy WF, McFersen JR (1998) Microsatellite (SSR) markers reveal genetic identities, genetic diversity and relationships in a *Malus* x *domestica* borkh. core collection. Theor Appl Genet 97:671–683

Huelsenbeck JP, Andolfatto P (2007) Inference of population structure under a Dirichlet process model. Genetics 175:1787–1802

Kitahara K, Matsumoto S, Yamamoto T, Soejima J, Kimura T, Komatsu H, Abe K (2005) Molecular characterization of apple cultivars in Japan by S-RNase analysis and SSR markers. J Amer Soc Hort Sci 130:885–892

Lamboy WF, Yu J, Forsline PL, Weeden NF (1996) Partitioning of allozyme diversity in wild populations of *Malus sieversii* L. and implications for germplasm collection. J Amer Soc Hort Sci 121:982–987

Lewis PO, Zaykin D (2001) GDA User's manual. Department of Ecology and Evolutionary Biology, University of Connecticut. 27 July 2004. <http://lewis.eeb.uconn.edu/lewishome>

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

Liebhard R, Kellerhals M, Pfammatter W, Jertmini M, Gessler C (2003) Mapping quantitative physiological traits in apple (*Malus x domestica* Borkh.). Plant Mol Biol 52:511–526

Luby J, Forsline P, Aldwinckle H, Bus V, Giebel M (2001) Silk road apples—collection, evaluation, and utilization of *Malus sieversii* from Central Asia. HortScience 36:225–231

Luby JJ, Alspach PA, Bus VGM, Oraguzie NC (2002) Field resistance to fire blight in a diverse apple (*Malus* sp.) germplasm collection. J Amer Soc Hort Sci 127:245–253

Morgan J, Richards A (1993) The book of apples. Ebury Press, London

Noiton DAM, Alspach PA (1996) Founding clones, inbreeding, coancestry, and status number of modern apple cultivars. J Amer Soc Hort Sci 121:773–782

Nordborg M, Hu TT, Ishino Y, Jhaveri J, Toomajian C, Zheng H, Bakker E, Calabrese P, Gladstone J, Goyal R, Jakobsson M, Kim S, Morozov Y, Padhukasahasram B, Plagnol V, Rosenberg NA, Shah C, Wall JD, Wang J, Zhao K, Kalbfleisch T, Schulz V, Kreitman M, Bergelson J (2005) The pattern of polymorphism in *Arabidopsis thaliana*. PLoS Biology 3:1289–1299

Oraguzie NC, Yamamoto T, Soejima J, Suzuki T, DeSilva HN (2005) DNA fingerprinting of apple (*Malus* spp.) rootstocks using simple sequence repeats. Plant Breed 124:197–202

Pereira-Lorenzo S, Ramos-Cabrer AMR, Ascasibar-Errasti J (2003) Analysis of apple germplasm in northwestern Spain. J Amer Soc Hort Sci 128:67–84

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

Ramos-Cabrer AM, Diaz-Hernandez MB, Pereira-Lorenzo S (2007) Morphology and microsatellites in Spanish apple collections. J Hort Sci Biotechnol 82:257–265

Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW (2002) Genetic structure of human populations. Science 298:2381–2385

Van Treuren R, van Soest LJM, van Hintum ThJL (2001) Marker-assisted rationalization of genetic resource collections: a case study with flax using AFLPs. Theor Appl Genet 103:144–152

Vavilov NI (1992) Origin and geography of cultivated plants. D. Love (translator). Cambridge University Press, Cambridge, England

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

Volk GM, Richards CM, Reilley AA, Henk AD, Forsline PL, Aldwinckle HS (2008) Genetic diversity and disease resistance of wild *Malus orientalis* from Turkey and southern Russia. J Am Soc Hort Sci 133:383–389

Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. Mol Ecol 15:1419–1439

Watkins R (1995) Apple and pear. In: Smartt and Simmonds (eds) Evolution of crop plants. Longman scientific and technical, Essex, England, pp 418–422

Yan G, Hong L, Song W, Chen R (2008) Genetic polymorphism of *Malus sieversii* populations in Xinjiang, China. Genet Res Crop Evol 55:171–181

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

Zhou ZQ, Li YN (2000) The RAPD evidence for the phylogenetic relationship of the closely related species of cultivated apple. Genet Res Crop Evol 47:353–357