Genetic Diversity and Characterization of a Core Collection of *Malus* Germplasm Using Simple Sequence Repeats (SSRs)

Sarah M. Potts · Yuepeng Han · M. Awais Khan · Mosbah M. Kushad · A. Lane Rayburn · Schuyler S. Korban

Published online: 28 December 2011 © Springer-Verlag 2011

Abstract Simple sequence repeats (SSRs) were used to assess genetic diversity and study genetic relatedness in a large collection of *Malus* germplasm. A total of 164 accessions from the *Malus* core collection, maintained at the University of Illinois, were genotyped using apple SSR markers. Each of the accessions was genotyped using a single robust SSR marker from each of the 17 different linkage groups in *Malus*. Data were subjected to principal component analysis, and a dendrogram was constructed to establish genetic relatedness. As expected, this diverse core collection showed high allelic diversity; moreover, this allelic diversity was higher than that previously reported. Cluster analysis revealed the presence of four distinct clusters of accessions in this collection.

Keywords Apple · Microsatellite markers · Genetic diversity · Genetic relatedness · Germplasm characterization

S. M. Potts · M. M. Kushad · A. L. Rayburn · S. S. Korban (⊠) Department of Crop Sciences, University of Illinois at Urbana–Champaign, Urbana, IL 61801, USA
e-mail: korban@illinois.edu

Y. Han

Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Wuhan Botanical Garden, Chinese Academy of Sciences, Moshan, Wuhan 430074, People's Republic of China

M. A. Khan S. S. Korban Department of Natural Resources and Environmental Sciences, University of Illinois at Urbana–Champaign, Urbana, IL 61801, USA

Introduction

Availability of diverse *Malus* germplasm is critical for pursuing successful apple breeding efforts, as it increases genetic diversity and allows for development of new apple cultivars with enhanced and/or desirable traits. This also aids in diversifying the gene pool and preserving those unique genetic traits available in this material.

When characterizing plant germplasm, incidence of multiple clones of genetic material as well as mislabeling of accessions may occur, which are both costly and undesirable (Garkava-Gustavsson et al. 2008). Moreover, proper identification and characterization of plant germplasm will protect intellectual property as well as aid in identifying parents carrying genes of interest for breeding efforts (Goulão et al. 2001; Dávila et al. 1998). By selecting diverse parents and increasing genetic diversity through germplasm collections, progress can be made in apple plant breeding efforts towards developing new cultivars with economically valuable traits including those with enhanced fruit quality and disease and pest resistance.

As in vivo maintenance and management of *Malus* germplasm are labor-intensive, costly, and require commitment of land resources for germplasm conservation efforts, determining genetic identity and genetic relatedness among accessions also impacts efficiency and utilization of such germplasm collections in breeding programs (Kresovich and McFerson 1992; Russell et al. 1997). The constraints of management of the *Malus* germplasm collection have led to development of strategies for germplasm evaluation. One of these strategies is the development of a core collection consisting of accessions having high levels of genetic diversity that could serve as representatives of the entire genetic diversity available within a collection (Frankel 1984; Brown 1989; Marshall 1990; Brown 1995). Developing core subsets of a population enhances the efficiency of screening and evaluation of desirable target traits. To increase the usefulness of core collections, the genetic information must be clearly identified and documented (Hokanson et al. 1998).

The use of DNA-based molecular markers has been highly critical and valuable for pursuing studies to assess genetic diversity, determine genetic relatedness, and identify genes of interest (Han and Korban 2010; Chen et al. 2010; Baraket et al. 2011). Markers such as simple sequence repeats (SSRs) are highly polymorphic, reproducible, and are distributed throughout the genome, rendering these markers ideal for pursuing genetic diversity studies. This has proven successful for various other fruit crops such as apricot (Wang et al. 2011), cherry (Clarke and Tobutt 2009), and strawberry (Govan et al. 2008), among others. Therefore, the use of DNA-based markers to characterize the *Malus* germplasm is a highly valuable and reliable tool (Garkava-Gustavsson et al. 2008; Zhang et al. 2011).

Many apple SSR markers have been developed, and most have proven highly valuable for identifying and characterizing limited numbers of apple germplasm, identifying loci associated with target genes, and for map-based cloning efforts (Liebhard et al. 2002; Naik et al. 2006; Gasic et al. 2009; Han and Korban 2010; Zhang et al. 2011). The overall goal of this study is to assess genetic relatedness and diversity in a large collection of *Malus* germplasm, selected by the US Apple Crop Germplasm Committee as a core collection based on known and/or reported phenotypic and genetic traits, using SSR markers.

Materials and Methods

Plant Material and DNA Extraction

Young expanding leaves were collected from 164 *Malus* accessions maintained within the *Malus* core collection available at the University of Illinois (Table 1). This core collection, selected by the US Apple Crop Germplasm Committee based on either known or reported phenotypic and genotypic traits, is a subset of *Malus* accessions representing the genetic diversity of the entire *Malus* germplasm available at the clonal repository of the Plant Genetic Resources Unit in Geneva, NY, USA. It is comprised of wild accessions, old and new cultivars, as well as advanced selections.

DNA was extracted from leaf tissues of all of the above accessions following the protocol described by Kobayashi et al. (1998), but with a slight modification by extending the incubation period for an additional 20 min at 65°C. Moreover, after adding isopropanol, microfuge tubes were kept at -20° C for no less than 4 h.

 Table 1
 A listing of 164 Malus accessions along with their corresponding repository identities used in this study

Name	GMAL	PI
M. angustifolia	2349	589763
M. asiatica (PI 589869)	2711	589869
M. asiatica (PI 594099)	1879	594099
M. baccata Flexillis	1605	437055
M. baccata Hansen's #2	2477	589838
M. baccata jackii	2460	594110
M. baccata Mandshurica 2330	35	322713
M. baccata Rockii	423	588960
M. bhutanica CH97 03-02	4370	_
M. bhutanica prunifolia var. macrocarpa	259	588930
M. coronaria	2892	589976
M. florentina	185	588868
M. florentina Skopje P2	-	589385
M. fusca (PI 589933)	2837	589933
M. fusca (PI 589941)	2848	589941
M. fusca (PI 589975)	2891	589975
M. halliana (PI 589972)	2887	589972
M. halliana (PI 594112)	3143	594112
M. hartwigii	1866	589420
M. honanensis	3238	594113
M. hupehensis CH97 04-13	-	633812
M. hupehensis CH97 07-07	4427	_
M. hybrid Demir	202	588883
M. hybrid E11-24	2089	589571
M. hybrid E29-56	3195	590071
M. hybrid E31-10	3196	590072
M. hybrid E36-7	2088	589570
M. hybrid Kansas K14	108	588804
M. hybrid Prairifire	2453	589820
M. hybrid PRI 1176-1	3209	590085
M. hybrid PRI 1316	2407	589776
M. hybrid PRI 1346-2	2416	589785
M. hybrid PRI 1484-1	2421	589790
M. hybrid PRI 1744-1	2420	589789
M. hybrid PRI 1754-2	2425	589794
M. hybrid PRI 1773-6	2440	589807
M. hybrid PRI 1850-4	2423	589792
M. hybrid PRI 1918-1	2408	589777
M. hybrid PRI 2050-2	2452	589819
M. hybrid PRI 2377-1	2445	589812
M. hybrid PRI 2382-1	2406	589775
M. hybrid PRI 2482-100	2426	589795
M. hybrid PRI 333-9	2467	589829
M. hybrid PRI 77-1	2417	589786
M. hybrid Prima	1064	589181
M. hybrid Robart's Crab	1611	437057
M. hybrid White Angel	494	588992
M. ioensis	2939	590015
M. ioensis Bechtel crab	493	588991
M. ioensis Texana		596279

Table 1 (continued)

Table I (continued)	Table 1	(continued)
---------------------	---------	-------------

Table 1 (continued)			Table 1 (continued)		
Name	GMAL	PI	Name	GMAL	PI
M. kansuensis	1875	594097	M. transitoria (PI 589384)	1822	589384
M. kirghisorum (PI 589380)	1750	589380	M. transitoria (PI 589422)	1869	589422
M. kirghisorum (PI 590043)	3158	590043	M.×arnoldiana Arnold Crab	1220	589222
M. micromalus (PI 589955)	2865	589955	M.×dawsoniana	6	483254
M. micromalus (PI 594092)	273	594092	M.×domestica Anna	85	280400
M. micromalus (PI 594096)	1497	594096	M.×domestica Antonovka 1.5 pounds	2461	107196
M. orientalis (GMAL 4556.p)	4556.p	-	M.×domestica Antonovka 172670-B	2866	589956
M. orientalis (PI 594095)	1461	594095	M.×domestica Antonovka Kamenichka	498	588995
M. orientalis 99TU-08-02	4513	_	M.×domestica Brite Gold	2308	589726
M. orientalis 99TU-16-01	4535	_	M.×domestica Burgundy	150	588835
M. orientalis 99TU-20-01	4539	_	M.×domestica Chisel Jersey	112	588806
M. orientalis RUS 98 02-01	4467	_	M.×domestica Cortland	163	588848
M. orientalis RUS 98 03-05	_	612381	M.×domestica Crimson Beauty	566	589024
M. orientalis RUS 98 07-01	_	612385	M.×domestica Dorsett Golden	2804	589913
M. prunifolia 19651	2449	589816	$M. \times domestica$ Ein Shemer	109	280401
M. prunifolia Inuringo	2175	594103	$M. \times domestica$ Empire	157	588842
M. prunifolia microcarpa	2457	594109	M.×domestica Florina	26	588747
M. prunifolia PRI 384-1	2827	589930	M.×domestica Fuji Red sport type 2	_	588844
M. prunifolia Xanthocarpa	2470	589832	$M. \times domestica$ Gala	1730	392303
M. numila	3163	323617	$M \times domestica$ Golden Delicious	3490	590184
M. rockii	1867	589421	$M. \times domestica$ Granny Smith	199	588880
M. sieversii (GMAL 4198.a)	4198.a	_	<i>M</i> × <i>domestica</i> Gravenstein Washington Red	152	588837
M sieversii (GMAL 4256 d)	4256 d	_	$M \times domestica$ Haralson	1982	589469
M sieversii (PI 596280 a)	_	596280 a	$M \times domestica$ Idared	156	588841
M sieversii (PI 596282 a)	_	596282 a	M × domestica Ingol	1953	589441
M sieversii (PI 596283 a)	_	596283 a	$M \times domestica$ Irish Peach	1115	104727
M. sieversii FORM 181(35-01)	_	613969	M × domestica James Grieve	1043	246464
M sieversii KAZ 93-24-01	3554	_	(Red Rosamund Strain)	1015	210101
M. sieversii KA7 93-42-01	3574	_	M.×domestica Jonafree	2872	589962
M sieversii KAZ 95 18-02P-33	5574	633801	M.×domestica Keepsake	2782	589894
M. sieversii KAZ 96 01-01P-20		633922	M.×domestica Kimball McIntosh 2-4-4-4	859	589122
M. sieversii KAZ 96 06-01P	_	599805	M.×domestica Koningszuur	1139	188517
$M_{\rm signarcii}$ KA7 96 07-04	_	613002	M.×domestica Korichnoe Polosatoje	2005	589491
M signarcii KAZ 96 07-04		613004	M.×domestica Lady	664	589053
M. sieversii KAZ 96 07-00	-	613058	M.×domestica Liberty	284	588943
M. sieversii KAZ 96 07-07		613008	M.×domestica Marshall McIntosh	508	588998
M. sieversii KAZ 90 08-10 M sieversii KAZ 96 08 17	-	613000	M.×domestica Medaille d'Or	2223	594108
$M_{\rm circuratii} KAZ 90 00-17$		614000	M.×domestica Mollie's Delicious	471	588981
M. sieversii KAZ 96 09-02	-	633020	M.×domestica Monroe	73	588772
M. sieversii KAZ 96 09-05 (PI 633920)1		633020	M.×domestica Murray	2000	589486
M. sikkimongia	1929	580300	M.×domestica Northern Spy	190	588872
M. sustabilis	1020	504100	M.×domestica Nova Easygro	153	588838
M. specialitis	262	260855	M.×domestica Novosibirski Sweet	1992	589478
M = sylvestris (PI 520323)	1820	580387	M.×domestica Petrel	2883	589970
$\frac{1}{1} \cdot \frac{1}{2} \cdot \frac{1}$	1620	J07302	M.×domestica Poeltsamaa Winter Apple	1607	383515
M. sylvestris Hartmann Muhla 1 y Okamuant - 2	2324	632027	M.×domestica Redfree	2875	594111
<i>M. sylvestris</i> Hartmann-Munie 1×Oberwartha 2	-	033827	M.×domestica Reinette Simirenko	2025	483257
<i>M. sylvestris</i> Oberwartna 5x Klipphausen	4495	-	M.×domestica Rhode Island Greening	2035	589520
<i>M. sylvestris</i> Oelsen 2X Hartmann Mahlel	-	633826	<i>M.×domestica</i> Rome Beauty Law	165	588850
M. toringo MA #4	2868	589958	M.×domestica Rosemary Russet	2180	589648
<i>M. toringo</i> Steboldii (PI 589/49)	2333	589749	$M. \times domestica$ Spokane Beauty	532	589006
M. toringo Sieboldii (PI 594094)	365	594094			

Table 1 (continued)

Name	GMAL	PI
M.×domestica Sweet Delicious	417	588955
M.×domestica Trent	2004	589490
M.×domestica Viking	1946	589434
M.×domestica Virginia Gold	80	588778
M.×domestica Wijcik McIntosh	3492	590186
M.×domestica Winter Majetin	2176	589645
M.×magdeburgensis	422	588959
M.×robusta Persicifolia	1821	589383
M.×soulardii	1829	589391
M.×sublobata Yellow Autumn Crab	250	588922
M. yunnanensis Vilmorin	537	271831
M. zhaojiaoensis CH97 06-2	_	633816
^a Unlabeled	_	_

^a Unlabeled accession was determined to be genetically identical to *M*. hybrid "Kansas K14"

PCR Amplification and Capillary Electrophoresis

DNA fragments were subjected to PCR amplification using 17 robust SSR markers spanning all 17 linkage groups (LG) of *Malus*. These SSR markers included the following: Hi02C07, CH02C06, GD12, NZ05g8, CH05f06, CH03d07, CH04e05, CH01h10, CH01f03b, CH02c11, CH02d08, CH01f02, GD147, CH04c07, CH02c09, CH04f10, and CH01h01 (Table 2). These markers were previously evaluated and developed by the European Cooperative Program for Plant Genetic Resources and were found to be highly polymorphic

(Evans et al. 2007). PCR reactions were performed in 96well plates in a total volume of 10 μ L containing 50 ng template DNA, 4.25 nuclease-free water, 0.3 mM MgCl₂, 0.2× Green Go*Taq*[®] Flexi Buffer, 0.05 U Go*Taq*[®] DNA polymerase (Promega, Madison, WI, USA), 0.04 mM of each dNTP, 0.25 μ L forward primer (10 μ M), 0.25 μ L reverse primer (10 μ M), and 0.15 μ L M13 fluorescent dye (10 μ M).

PCR amplification was carried out using either a Thermo Fisher Scientific multi-block thermal cycler (Pittsburgh, PA, USA) or an MJ Research PTC-100 or PTC-200 (Ramsey, MN, USA). Amplifications were performed using the following conditions: initial denaturation at 94°C for 4 min, 5 cycles of 94° C for 1 min, 54°C for 1 min, 72°C for 1 min, 30 cycles of 94°C for 1 min, 52°C for 1 min, 72°C for 1 min, and a final extension at 72°C for 30 min. This was followed by holding at 4°C.

Individual PCR products were labeled with one of four M13 dyes, FAM, VIC, PET, and NED, along with a corresponding LIZ 600 size standard. PCR products were then pooled for electrophoresis. Amplified PCR products were separated at the W.M. Keck Center at the University of Illinois using an ABI 3730xl sequencer (Applied Biosystems, Inc., Foster City, CA, USA). Raw fragment size data were analyzed using GeneMapper[™] Software ver. 4.0 (Applied Biosystems, Inc.), and all automated results were manually reviewed.

Statistical Analysis

Statistical analysis was performed using SAS[®] 9.2 software. The proc univariate was used to determine data normality. Proc corr was used to assess fragment length data correlations.

Table 2 SSR marker information for all 17 17 robust SSR markers used 17	Linkage group	Source ^z	Locus	Locus type	Number of alleles	Expected heterozygosity ^y	Range of fragment length (bp)
for genotyping the <i>Malus</i> germplasm	1	4	Hi02C07#	Pres-multi	5	nd	108–149
collection used in	2	3	CH02C06+	Single	8	0.85	216-254
this study	3	2	GD12*	Unknown	12	0.758	141–191
	4	1	NZ05g8~	Single	6	0.76	115-147
	5	3	CH05f06+	Single	5	0.74	166–184
	6	3	CH03d07+	Single	8	0.8	186–226
	7	3	CH04e05+	Pres-multi	8	nd	174–227
	8	3	CH01h10+	Single	5	0.65	94–114
	9	3	CH01f03b+	Single	7	0.8	139–183
	10	3	CH02c11+	Single	7	0.78	219–239
	11	3	CH02d08+	Single	7	0.82	210-254
² Sources of these SSR markers include the following: 1) Guilford et al. (1997); 2) Hokanson et al. (1998); 3) Liebhard et al. (2002); and 4) Silfverberg-Dilworth et al. (2006) ^Y nd, not detected	12	3	CH01f02+	Single	7	0.79	174-206
	13	2	GD147*	Single	6	nd	135-155
	14	3	CH04c07+	Single	8	0.82	98–135
	15	3	CH02c09+	Single	6	0.77	233-257
	16	3	CH04f10+	Single	9	0.88	144–25
	17	3	CH01h01+	Single	6	0.8	114–134

Proc cluster and proc tree were used to create the dendrogram using options rsquare and simple for proc cluster, and options ftext=triplexu, hsize=10, ftext=0.2, vsize=10, interval=0.2, htext=0.2, horizontal, vpages=5, height=rsq, and inc=1.4 for proc tree. Proc princomp was used to obtain principal components, and proc g3d was used to create a scatter plot of the accessions with the options reset=all border, tilt=50, rotate= 30, color=color, and shape=shape.

Data for each SSR were assessed using the correlation procedure to determine correlations among these variables. Ward's minimum variance method, a non-hierarchical cluster method, was used to generate clusters. The number of clusters was determined by the cubic clustering criterion, the pseudo-F, and the pseudo- T^2 along with R^2 values. Data were then subjected to a principal component analysis to create principal components (PCs) for the construction of a three-dimensional scatter plot (Johnson 1998).

Expected heterozygosity and observed heterozygosity were calculated using the program GENEPOP (Raymond and Rousset 1995; Rousset 2008). Polymorphic information content (PIC) (Botstein et al. 1980; Hearne et al. 1992) was calculated using the program Cervus (Kalinowski et al. 2007). Effective alleles per locus were calculated according to Morgante et al. (1994).

Of 17 primer pairs, 10 were selected for final analysis on the basis of reliable amplification and signal strength. These included the following markers: Hi02c07, GD147, CH04e05, CH04c07, CH03d07, CH02c09, CH01h10, CH04f10, CH02d08, and CH01f03b. This number of markers was similar to that reported in previous studies, which have used anywhere from eight markers (Benson et al. 2001; Király et al. 2009) to 14 markers (Guilford et al. 1997).

Results and Discussion

The *Malus* core collection used in this study was originally selected as the best representative of available phenotypic and

831

genetic diversity for various economic traits of interest for use in evaluations for various biotic and abiotic stresses as well as fruit quality traits (Forsline 1996).

All SSR primer pairs used in this study generated multiple fragments in the *Malus* germplasm core collection. Of 164 *Malus* accessions, 39 did not amplify at least one PCR product. Due to the nature of multivariate analysis, accessions with missing data were not used for either the dendrogram or the scatter plot. A total of 125 accessions were included in generating both the dendrogram and the scatter plot (Table 3). However, due to the nature of the analysis and the strength of the remaining data, all 164 accessions were included in calculating expected heterozygosity, observed heterozygosity, PIC, and effective alleles per locus.

A total of 283 fragments were amplified using the selected 10 SSR markers (Table 4). In general, each primer pair amplified several alleles in large numbers, while many alleles were either rare or unique, representing only 2–5% of the entire *Malus* core collection. Overall, a higher allelic diversity was observed in this *Malus* core collection than that reported in previous studies (Hokanson et al. 1998; Kitahara et al. 2005; Garkava-Gustavsson et al. 2008; Zhang et al. 2011). Discrepancies in allelic diversity could be attributed to differences in DNA isolation protocols and perhaps increased ability to detect 1–2 bp length differences utilizing modern fragment analysis technologies.

In this study, rare alleles (<5% of the total alleles per marker) have been detected at a frequency of 39% compared to frequencies of 53% (Garkava-Gustavsson et al. 2008) and 59% (Hokanson et al. 1998). These discrepancies could be attributed to the use of different SSRs, as only three markers are common among these three different studies. Another possible explanation is that rare alleles have been selected for in domesticated cultivars, whereas studies targeting wild germplasm are not selected for these rare alleles. Rare alleles are critical for maintaining genetic diversity as they are unique and they are likely

Table 3Genetic diversityinformation of theMalusgermplasm collectionused in this study as revealedby 10 robust SSR markersused for genotyping

Linkage group	Locus	Expected heterozygosity	Observed heterozygosity	Polymorphic information content	Effective number of alleles
1	Hi02C07#	0.857	0.752	0.8399	6.863
6	CH03d07+	0.920	0.795	0.9122	12.096
7	CH04e05+	0.865	0.679	0.8516	7.252
8	CH01h10+	0.821	0.632	0.8023	5.507
9	CH01f03b+	0.874	0.756	0.8596	7.784
11	CH02d08+	0.871	0.745	0.8583	7.586
13	GD147*	0.901	0.753	0.8899	9.816
14	CH04c07+	0.899	0.690	0.8873	9.598
15	CH02c09+	0.911	0.745	0.9011	10.919
16	CH04f10+	0.954	0.600	0.9492	20.489

Cluster

Table 4 A listing of 125 Malus accessions amplified by the mostinformative 10 SSRs and by designated cluster

 Table 4 (continued)

informative 10 SSRs and by designated cluster		Name	
Name	Cluster		
M bhutanica prunifolia yar macrocarpa	1	M. orientalis (PI 594095) M. orientalis RUS 98 03-05	
M. bybrid F29-56	1	M. orientalis ROS 90 05 05	
M. hybrid E21-10	1	M. siquarsii (PI 596282 a)	
M. orientalis 99TU-16-01	1	M, sieversii (PI 596283 a)	
M. orientalis PUS 98 02 01	1	M. sieversii (11390203.a) M. sieversii V A 7 02 42 01	
M. prunifolia DDI 284 1	1	M. sieversii KAZ 95-42-01	
M. prunyoliu FKI 564-1	1	M. subversul KAZ 90 00-01F	
M. sieversii (F1 590280.a)	1	M. sylvestris (PI 619168)	
<i>W. steversti</i> FORM 181(33-01)	1	M. sylvestris Oelsen 2X Hartmann Manlei	
<i>A. steversti</i> KAZ 90 07-00	1	M.× aawsoniana	
<i>1. sieversii</i> KAZ 96 07-07	1	M.×domestica Anna	
1. sieversii KAZ 96 08-16	1	M.×domestica Ein Shemer	
1. sieversii KAZ 96 09-05 (PI 633920)1	1	M.×domestica Empire	
1. sylvestris (PI 589382)	1	<i>M.×domestica</i> Fuji Red sport type 2	
<i>I.×arnoldiana</i> Arnold Crab	1	M.×domestica Haralson	
1.×domestica Antonovka 1.5 pounds	1	M.×domestica James Grieve (Red Rosamund S	
<i>I.×domestica</i> Burgundy	1	<i>M.×domestica</i> Redfree	
<i>I.×domestica</i> Chisel Jersey	1	<i>M.×domestica</i> Sweet Delicious	
1.×domestica Cortland	1	M.×domestica Virginia Gold	
1.×domestica Dorsett Golden	1	M. bhutanica CH97 03-02	
1.×domestica Gravenstein Washington Re	1	M. hupehensis CH97 04-13	
I.×domestica Irish Peach	1	M. hybrid E36-7	
I.×domestica Kimball McIntosh 2-4-4-4	1	M. hybrid Prairefire	
!.×domestica Koningszuur	1	M. hybrid PRI 1346-2	
.×domestica Korichnoe Polosatoje	1	M. hybrid PRI 1850-4	
.×domestica Marshall McIntosh	1	M. hybrid PRI 2050-2	
L×domestica Murray	1	M. kirghisorum (PI 590043)	
I.×domestica Nova Easygro	1	M. orientalis RUS 98 07-01	
<i>I.×domestica</i> Petrel	1	M. prunifolia microcarpa	
I.×domestica Poeltsamaa Winter Apple	1	M. sieversii (GMAL 4198.a)	
1.×domestica Rosemary Russet	1	M. sieversii (GMAL 4256.d)	
I.×domestica Spokane Beauty	1	M. sieversii KAZ 93-24-01	
I.×domestica Viking	1	M. sieversii KAZ 96 01-01P-20	
I.×domestica Wijcik McIntosh	1	M. sieversii KAZ 96 07-04	
<i>I.×soulardii</i>	1	M. sieversii KAZ 96 08-17	
Unlabeled	2	M.×domestica Florina	
1. angustifolia	2	M.×domestica Gala	
<i>I. asiatica</i> (PI 589869)	2	M.× domestica Golden Delicious	
1. asiatica (PI 594099)	2	$M. \times domestica$ Idared	
I. florentina	2	$M. \times domestica$ Lady	
<i>1. hybrid</i> Demir	2	<i>M. baccata</i> Rockii	
I. hvbrid Kansas K14	2	M. honanensis	
I. hvbrid PRI 1484-1	- 2	M. hvbrid PRI 1744-1	
1 hybrid PRI 1754-2	- 2	M hybrid PRI 77-1	
A hybrid PRI 1773-6	2	M hybrid White Angel	
A hybrid PRI 1918-1	2	M. micromalus (PI 59/002)	
/ hybrid PRI 2482-100	2	M micromalus (PI 594092)	
1. kirahisorum (PI 580380)	2	M. mutomuus (11 574070)	
. паднизонин (11 307300) (miaromatus (DI 590055)	2	M. pranijona Aantiiocarpa	
1. micromatus (P1 389933)	L	IVI. FOCKII	

 Table 4 (continued)

Name	Cluster
M. toringo Sieboldii (PI 589749)	4
M. toringo Sieboldii (PI 594094)	4
M.×domestica Novosibirski Sweet	4
M.×sublobata Yellow Autumn Crab	4
M. zhaojiaoensis CH97 06-2	4
M. baccata Hansen's #2	5
M. baccata jackii	5
M. halliana (PI 589972)	5
M. hartwigii	5
M. hybrid PRI 333-9	5
M. hybrid Robart's Crab	5
M. prunifolia Inuringo	5
M. spectabilis	5
M. sylvestris Oberwartha 5x Klipphausen	5
M. transitoria (PI 589384)	5
M.×domestica Jonafree	5
M.×domestica Keepsake	5
M.×domestica Monroe	5
M.×domestica Winter Majetin	5
M. hybrid PRI 1316	6
M. hybrid PRI 2382-1	6
M. hybrid Prima	6
M. orientalis 99TU-20-01	6
M. sieversii KAZ 96 09-05 (PI 633920)2	6
M.×domestica Ingol	6
M.×domestica Northern Spy	6
M.×domestica Rhode Island Greening	6
M.×domestica Rome Beauty Law	6

to be involved in plant adaptation to environmental shifts (Richter et al. 1994; Bengtsson et al. 1995). Thus, these alleles are important components of this core collection.

In this study, 164 *Malus* accessions from diverse genetic backgrounds include *Malus* species, $M. \times domestica$ cultivars, and selections, while previous studies have focused on assessing allelism using smaller collections consisting primarily of $M. \times domestica$ cultivars. As modern cultivars are derived from a relatively narrow genetic base, it is not unexpected that these collections would be less genetically diverse. Only five founding clones were progenitors for 64% of a total of 439 cultivars in a co-ancestry study (Noiton and Alspach 1996), thus pointing to the highly shared genetic identity of modern apple cultivars.

The cluster function of SAS 9.2 (proc cluster) produced a dendrogram composed of six distinct clusters (Fig. 1). These clusters were populated as follows: The first cluster consisted of 34 accessions, a second cluster consisted of 33 accessions, a third cluster consisted of 21 accessions, and a fourth cluster

consisted of 14 accessions, while both fifth and sixth clusters consisted of 14 and nine accessions, respectively, for a total of 125 accessions. In addition, a scatter plot of these accessions (Fig. 2) was produced using the first, second, and third principal components, and accounting for 35% of the variance. Principal component one was comprised mainly of SSR markers CH01h10, CH03d07, GD147, and CH01f03b. Principal component two was mainly comprised of markers CH02c09, CH04e05, and CH01f03b1, whereas principal component 3 was mainly comprised of markers CH04e05, CH01f03b, and CH04f10.

The expected heterozygosity ranged from 0.821 to 0.954, with a mean value of 0.887, while observed heterozygosity ranged from 0.600 to 0.795, with a mean value of 0.715 (Table 2). PIC values ranged from 0.802 to 0.949 with a mean value of 0.875 (Table 2). The range of effective alleles per locus was wide, ranging from 5.507 to 20.489, with an average of 9.791 (Table 2).

Similarities were found both by pedigree analysis and by previously reported genetic relatedness studies. As expected, "McIntosh" sports "Kimball McIntosh," "Marshall McIntosh," and "Wijcik McIntosh" were genetically identical for all markers, as similarly reported by Hokanson et al. (1998). "Golden Delicious" is a parent of "Gala" (Kouassi et al. 2009) and both closely clustered together.

Multiple groups, including "McIntosh" sports and "Cortland," "Ein Shemer" and "Virginia Gold," "Koningszuur" and "Spokane Beauty," "Murray" and "Viking," as well as "Northern Spy" and "Rhode Island Greening" are clustered together in this study, and this is similar to findings reported by Hokanson et al. (1998). "PRI 1484-1," "PRI 1773-6," and "E36-7" are clustered at high proximity to each other, and this is similar to findings of Hokanson et al. (2001). Moreover, accessions "PRI 1918-1" and "PRI 2482-100," "PRI 2050-2," and "PRI 1346-2," as well as "Demir" and *M. kirghisorum* clustered similarly to those reported by Hokanson et al. (2001).

Previously, Hokanson et al. (1998) have reported that "Murray" is distantly clustered from other "McIntosh" accessions. However, in this study, "Murray" is clustered in close proximity to other "McIntosh" accessions, which is to be expected as "Murray" is to known to be a "McIntosh" seedling. In other findings in this study, "Ein Shemer" is separated from both "Golden Delicious" and "Gala"; "Northern Spy" is separated from "Jonafree"; "Irish Peach" is separated from "Keepsake"; and "Korichnoe Polosatoje" is separated from "Winter Majetin." Previously, Hokanson et al. (1998) have reported instead that the above groups of accessions are clustered together. Among other discrepancies, M.×soulardii and "Arnold Crab" are found in different clusters by Hokanson et al. (2001), while these have clustered similarly in this study. This is also the case with "Hansen's #2" and Malus halliana (PI 589972), "Novosibirski Sweet" and

Fig. 1 A dendrogram of 125 *Malus* accessions clustered based on genotypic analysis of 10 robust SSR markers into six distinct clusters





Malus accessions based on 10 SSR markers

Fig. 2 A scatter plot of 125 *Malus* accessions analyzed based on principal components, accounting for 35% of variability detected in this material. Principal component one (Prin1) was comprised mainly of SSR markers CH01h10, CH03d07, GD147, and CH01f03b. Principal component two

(Prin2) was mainly comprised of markers CH02c09, CH04e05, and CH01f03b1, while principal component 3 (Prin3) was mainly comprised of markers CH04e05, CH01f03b, and CH04f10

"White Angel," as well as "Robert's Crab" and "Inuringo." Such discrepancies could be due to use of either different or possible mislabeling in one of these collections. A solution to mislabeling has been proposed via the use of common DNA samples from a single source as control across different studies, which could then be used to check for trueness-to-type (Evans et al. 2009).

Another possible explanation for the above observed differences could be attributed to the different statistical analyses used in these two studies. While the unweighted pair-group method has been used by Hokanson et al. (1998), Ward's minimum variance method is used in this study. The unweighted pair-group method is a hierarchical method of creating clusters, which results in a nested dendrogram. Additionally, the unweighted pair-group method assumes equal rates of evolution between lineages, whereas Ward's minimum variance method is nonhierarchical, and it is designed to obtain clusters with the least amount of within-group variance and the most amount of between-group variance, thus allowing for better separation of distinct groups. It is also likely that these observed differences may be attributed to differences in DNA separation methodologies used in these two studies. Hokanson et al. (1998) have used the ABI 377 or 373 DNA sequencing system (Applied Biosystems, Inc.) for gel-based separation, whereas the ABI 3730xl sequencer (Applied Biosystems, Inc.) for separation via capillary electrophoresis is used in this study. It is likely that capillary electrophoresis separation has resulted in different DNA fragment sizes than is possible with using gel-based separation.

Although there were clusters of accessions with common origin, or research station if the accession was a selection, a sizeable portion of the accessions did not cluster in a meaningful way. Decreased clustering by origin and taxonomy has been exhibited previously as a result of adding species and derived hybrids to smaller cultivar collections (Hokanson et al. 2001; Dunemann et al. 1994).

In this study, "Kansas K14" and an unlabeled accession (planted adjacent to "Kansas K14") were grouped together with an R^2 value of 1.00, thus suggesting that the unlabeled accession was in fact a duplicate clone of "Kansas K14." Additionally, two accessions, *M. sieversii* "KAZ 93-24-01" and *M. orientalis* "RUS 98 07-01," unexpectedly grouped together with an R^2 value of 1.00. These two accessions originated from different collections; thus, it is likely that these accessions are likely to have been mislabeled at either the original collection or subsequently in the core collection used in this study.

Taking into account "Kansas K14" and its duplicate unlabeled accession along with other duplicate pairs as well as those genetically identical "McIntosh" accessions, a total of 121 unique genotypes have been identified from 125 *Malus* accessions that were genetically characterized using the final set of markers. Using a set of ten robust SSR markers, it was possible to differentiate all accessions. All other additional markers were useful for confirmation of these findings. Moreover, two accessions with identical PI numbers (PI 633920 *M. sieversii* "KAZ 96 09-05"), but at different locations (K1-17-15 and K1-17-7), were not genetically similar as they were clustered in different groups, clusters 3 and 6, respectively.

Although six clusters generated by the scatter plot overlapped and were not clearly distinguishable, it should be noted that the scatter plot was a three-dimensional graphical representation of these data and accounted for only 35% of the existing variability. As only three of the PCs were used for visualization, it was assumed that clusters would be clearly differentiated in the ten-dimensional space required to observe most of the variability (Johnson 1998).

A relatively high allelic diversity has been reported in studies that included wild *Malus* species (Hokanson et al. 2001; Richards 2009; Zhang et al. 2011). Using a different set of SSRs on a slightly smaller number of similar accessions, Hokanson et al. (2001) have reported slightly higher numbers of effective alleles than that found in this study. In contrast, two previous studies evaluating only apple cultivars have identified significantly lower numbers of effective alleles (Hokanson et al. 1998; Garkava-Gustavsson et al. 2008). This is an expected finding due to lower levels of genetic diversity present in modern cultivars (Noiton and Alspach 1996).

Overall, levels of heterozygosity detected in this study were sometimes different than those reported previously. Although Hokanson et al. (1998, 2001) have used a similar collection, they reported lower levels of heterozygosity than that detected in this study. This could be attributed to the use of different marker sets, especially since markers used in this study had higher PIC values than those used in earlier studies (Table 2). In contrast, three studies focusing on apple cultivars (Liebhard et al. 2002; Kitahara et al. 2005; Garkava-Gustavsson et al. 2008) displayed higher levels of heterozygosity than that reported in this study. During cross-hybridization and selection efforts in crop improvement, higher levels of heterozygosity are generated (Lamboy and Alpha 1998). As the Malus core collection used in this study included wild Malus species, heterozygosity levels were lower than what would generally be observed in collections containing only domesticated apples.

As expected, this diverse core collection of *Malus* germplasm showed high allelic diversity. Although fewer rare alleles were found than in previous studies, 39% of the alleles detected in this collection were only present in five or fewer accessions, indicating alleles to maintain genetic diversity (Richter et al. 1994; Bengtsson et al. 1995). Genetic relatedness, as determined by cluster analysis, showed both similarities and dissimilarities to previously studies. One unlabeled accession was discovered to be a replicate of another accession, "Kansas K14," and two accessions thought to be duplicate were not genetically identical. The set of ten SSR markers were sufficient to differentiate all accessions in the core collection except for three "McIntosh" sport mutations, the determined "Kansas K14" replicate, and a duplicate accession. Heterozygosity was lower than studies focusing on $M. \times domestica$ cultivars, likely due to increased heterozygosity as a result of selection (Lamboy and Alpha 1998).

Acknowledgments This research was supported by Pioneer Hi-Bred International, Inc., Illinois Plant Breeding Center, USDA-NIFA-SCRI grant AG 2009-51181-06023, University of Illinois Office of Research Project 875-325, and University of Illinois Office of Research Project 875-922.

References

- Baraket G, Chatti K, Saddoud O, Ben Abdelkarim A, Mars M, Trifi M, Hannachi A (2011) Comparative assessment of SSR and AFLP markers for evaluation of genetic diversity and conservation of fig, *Ficus carica* L., genetic resources in Tunisia. Plant Mol Biol Rep 29:171–184
- Bengtsson B, Weibull P, Ghatnekar L (1995) The loss of alleles by sampling: a study of the common outbreeding grass *Festuca ovina* over three geographical scales. Hereditas 122:221–238
- Benson L, Lamboy W, Zimmerman R (2001) Molecular identification of Malus hupehensis (tea crabapple) accessions using simple sequence repeats. HortScience 36:961–966
- Botstein D, White R, Skolnick M, Davis R (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Human Genet 32:314–331
- Brown A (1989) The case for core collections. In: Brown AHD, Frankel OH, Marshall DR, Williams JT (eds) The use of plant genetic resources. Cambridge University Press, Cambridge, pp 136–156
- Brown A (1995) The core collection at the crossroads. In: Hodgkin T, Brown AHD, Hintum TJL, Morales EAV (eds) Core collections of plant genetic resources. Wiley, Chichester, pp 3–19
- Chen X, Xia Z, Fu Y, Lu C, Wang W (2010) Constructing a genetic linkage map using an F1 population of non-inbred parents in Cassava (*Manihot esculenta* Crantz). Plant Mol Biol Rep 28: 676–683
- Clarke J, Tobutt K (2009) A standard set of accessions, microsatellites and genotypes for harmonising the fingerprinting of cherry collections for the ECPGR. Acta Hort 814:615–618
- Dávila J, Sánchez de la Hoz M, Loarce Y, Ferrer E (1998) The use of random amplified microsatellite polymorphic DNA and coefficients of parentage to determine genetic relationships in barley. Genome 41:477–486
- Dunemann F, Kahnau R, Schmidt H (1994) Genetic relationships in Malus evaluated by RAPD 'fingerprinting' of cultivars and wild species. Plant Breed 113:150–159
- Evans K, Fernández F, Laurens F, Feugey L, Van de Weg E (2007) Harmonizing fingerprinting protocols to allow comparisons between germplasm collections. Eucarpia XII Fruit Selection Symposium, September 16–20, 2007, Zaragoza, Spain, pp 57–58
- Evans K, Fernández F, Govan C (2009) Harmonising fingerprinting protocols to allow comparisons between germplasm collections— *Pyrus*. Acta Hort 814:103–106

- Forsline P (1996) Core subsets in the USDA/NPGS with Apple as an example. Proceedings of the 2nd Workshop on Clonal Genetic Resources, Ottawa, Ontario, January 23–24, 1996. pp 172–175
- Frankel O (1984) Genetic perspectives of germplasm conservation. In: Arber WK, Llimensee K, Peacock WJ, Starlinger P (eds) Genetic manipulation: Impact on man and society. Cambridge University Press, Cambridge, pp 161–170
- Garkava-Gustavsson L, Brantestam A, Sehic J, Nybom H (2008) Molecular characterisation of indigenous Swedish apple cultivars based on SSR and S-allele analysis. Hereditas 145:99–112
- Gasic K, Han Y, Kertbundit S, Shulaev V, Iezzoni A, Stover E, Bell R, Wisniewski M, Korban SS (2009) Characteristics and transferability of new apple EST-derived SSRs to other Rosaceae species. Mol Breed 23:397–411
- Goulão L, Cabrita L, Oliveira C, Leitao J (2001) Comparing RAPD and AFLPTM analysis in discrimination and estimation of genetic similarities among apple (*Malus domestica* Borkh.) cultivars— RAPD and AFLP analysis of apples. Euphytica 119:259–270
- Govan C, Simpson D, Johnson A, Tobutt K, Sargent D (2008) A reliable multiplexed microsatellite set for genotyping *Fragaria* and its use in a survey of 60 *F*×*ananassa* cultivars. Mol Breed 22:649–661
- Guilford P, Prakash S, Zhu J, Rikkerink E, Gardiner S, Bassett H (1997) Microsatellites in *Malus xdomestica* (apple): abundance, polymorphism and cultivar identification. Theor Appl Genet 94:249–254
- Han Y, Korban SS (2010) Strategies for map-based cloning in apple. Crit Rev Plant Sci 29:265–284
- Hearne C, Ghosh S, Todd J (1992) Microsatellites for linkage analysis of genetic traits. Trends Genet 8:288–294
- Hokanson S, Szewc-McFadden A, Lamboy W, McFerson J (1998) Microsatellite (SSR) markers reveal genetic identities, genetic diversity and relationships in a *Malus xdomestica* Borkh. core subset collection. Theor Appl Genet 97:671–683
- Hokanson S, Lamboy W, Szewc-McFadden A, McFerson J (2001) Microsatellite (SSR) variation in a collection of *Malus* (apple) species and hybrids. Euphytica 118:281–294
- Johnson D (1998) Principal component analysis, and cluster analysis. In: Johnson DE (ed) Applied multivariate methods for data analysts. Brooks/Cole, Pacific Grove, pp 319–396
- Kalinowski S, Taper M, Marshall T (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol Ecol 16:1099–1106
- Király I, Tóth M, Pedryc A, Halász J, Deák T (2009) Parent identification of Hungarian apple cultivars using SSR markers. Acta Hort 839:471–477
- Kitahara K, Matsumoto S, Yamamoto T, Soejima J, Kimura T, Komatsu H (2005) Molecular characterization of apple cultivars in Japan by S-RNase analysis and SSR markers. J Amer Soc Hort Sci 130:885–892
- Kobayashi N, Horikoshi T, Katsuyama H, Handa T, Takayanagi K (1998) A simple and efficient DNA extraction method for plants, especially woody plants. Plant Tiss Cult Biotech 4:76–80

- Kouassi A, Durel C, Costa F, Tartarini S, van de Weg E, Evans K (2009) Estimation of genetic parameters and prediction of breeding values for apple fruit-quality traits using pedigreed plant material in Europe. Tree Genet Genomes 5:659–672
- Kresovich S, McFerson J (1992) Assessment and management of plant genetic diversity: considerations of intra- and inter-specific variation. In: Kresovich S (ed.) The impact of progress in genetics on plant resources (germplasm) conservation and utilization. Field Crops Res 29:185–204
- Lamboy W, Alpha C (1998) Using simple sequence repeats (SSRs) for DNA fingerprinting germplasm accessions of grape (*Vitis* L.) species. J Amer Soc Hort Sci 123:182–188
- Liebhard R, Gianfranceschi L, Koller B, Ryder C, Tarchini R, Van de Weg E (2002) Development and characterisation of 140 new microsatellites in apple (*Malus×domestica* Borkh.). Mol Breed 10:217–241
- Marshall D (1990) Crop genetic resources: current and emerging issues. In: Brown AHD, Clegg MH, Kahler AL, Weir BS (eds) Plant population genetics, breeding, and genetic resources. Wiley, Chichester, pp 367–388
- Morgante M, Rafalski A, Biddle P, Tingey S, Olivieri A (1994) Genetic mapping and variability of seven soybean simple sequence repeat loci. Genome 37:763–769
- Naik S, Hampson C, Gasic K, Bakkeren G, Korban S (2006) Development and linkage mapping of E-STS and RGA markers for functional gene homologues in apple. Genome 49:959–968
- Noiton D, Alspach P (1996) Founding clones, inbreeding, coancestry, and status number of modern apple cultivars. J Amer Soc Hort Sci 121:773–782
- Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J Heredity 86:248–249
- Richards C (2009) Genetic diversity and population structure in *Malus* sieversii, a wild progenitor species of domesticated apple. Tree Genet Genomes 5:339–347
- Richter T, Soltis P, Soltis D (1994) Genetic variation within and among populations of the narrow endemic, *Delphinium viridescens* (Ranunculaceae). Amer J Bot 81:1070–1076
- Rousset F (2008) Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. Mol Ecol Res 8:103–106
- Russell J, Fuller J, Macaulay M, Hatz B, Jahoor A, Powell W, Waugh R (1997) Direct comparison of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs. Theor Appl Genet 95:714–722
- Wang Y, Zhang J, Sun H, Ning N, Yang L (2011) Construction and evaluation of a primary core collection of apricot germplasm in China. Scient Hort 128:311–319
- Zhang Q, Li L, Zhao Y, Korban S, Han Y (2011) Evaluation of genetic diversity in Chinese wild apple species along with apple cultivars using SSR markers. Plant Mol Biol Rep. doi:10.1007/s11105-011-0366-6