Theor Appl Genet (1998) 97: 671–683

S. C. Hokanson · A. K. Szewc-McFadden
W. F. Lamboy · J. R. McFerson

Microsatellite (SSR) markers reveal genetic identities, genetic diversity and relationships in a *Malus* × *domestica* borkh. core subset collection

Received: 28 March 1998 / Accepted: 29 April 1998

Abstract A collection of 66 *Malus* × *domestica* Borkh. accessions from the USDA-ARS Plant Genetic Resources Unit's core collection was screened with a set of eight SSR (simple sequence repeat) primers developed at the PGRU in order to determine genetic identities, estimate genetic diversity, and to identify genetic relationships among these accessions. All eight primer pairs generated multiple fragments when used in amplification reactions with DNA from these accessions. High levels of variation were detected with a mean of 12.1 alleles per locus and a mean heterozygosity across all eight loci of 0.693. The eight primer pairs utilized in this study unambiguously differentiated all but seven pairs of accessions in this collection of 66 *M.* × *domestica* Borkh. genotypes. The probability of matching any two genotypes at all eight loci in this study was approximately 1 in 1 billion. The markers detected two misnamed accessions in the collection. Genetic-identity data produced a genetic-relatedness phenogram which was concordant with geographic origins and/or known pedigree information. These SSR markers show great promise as tools for managing *Malus* ex situ germplasm collections as well as for collection and preservation strategies concerning wild *Malus* populations in situ.

Communicated by P. M. A. Tigerstedt

S. C. Hokanson (✉)[1] · A. K. Szewc-McFadden ·
W. F. Lamboy · J. R. McFerson[2]
USDA-ARS Plant Genetic Resources Unit, Cornell University,
Geneva, NY 14456-0462, USA

Present addresses:
[1] USDA-ARS Fruit Lab, Building 010A, Rm. 210, BARC-W, 10300 Baltimore Ave., Beltsville, MD 20705-2350 USA
Tel.: (301) 504-6768
Fax: (301) 504-5062
E-mail: hokanson@asrr.arsusda.gov
[2] Washington Tree Fruit Research Commission,
1719 Springwater Ave., Wenatchee,
WA 98801, USA

Key words *Malus* × *domestica* Borkh. · Microsatellite markers · Genetic variation · Germplasm management · Core collections

Introduction

To be most efficiently managed and effectively utilized, germplasm collections must be well-characterized. A curator's ability to achieve this goal is often held captive by the triumvirate of rising costs, static budgets, and large collection sizes (Kresovich and McFerson 1992). Such considerations dictate that germplasm collections will be "leaner", exhibiting minimal redundancy with regard to genotypes, gene complexes or possibly even genes. Furthermore, to increase collection utility, information regarding the location of potentially useful genes and gene complexes within the collection and/or genome will need to be well-documented and readily available.

Two new trends in germplasm management are facilitating this characterization process. One is the development of core subsets, which are subsets of the germplasm collection at large. These subsets have been established to represent the genetic diversity found within the entire collection (Frankel 1984; Brown 1989 a; Marshall 1990; Brown 1995). Given large, and at times rapidly growing, collections, the establishment of core subsets facilitates systematic and rigorous characterizations of a more tractable set of genotypes within current budget and time constraints. Larger-scale evaluations can then be applied to the reserve collections with more precision and efficiency.

The second new trend involves the manner in which germplasm characterizations are being done. Historically, characterizations of germplasm collections have been carried out at several levels, from descriptions of taxonomy (Hilu 1989), biogeography (Lyman 1984; Nabhan 1985; Brush 1989; Zimmerer et al. 1991), and

morphology and agronomic characters (Chapman 1989) to biochemical analysis (Doebley 1989; Gepts 1990) and the study of molecular traits and markers (Clegg 1990; Gepts 1995). The emergence of new PCR-based molecular markers, such as randomly amplified polymorphic DNA (RAPDs), simple sequence repeats (SSRs), and amplified fragment length polymorphisms (AFLPs), has created the opportunity for fine-scale genetic characterizations of germplasm collections that were previously impossible. Since the PCR-based markers are highly polymorphic (however, see Broun and Tanksley 1996) and simple to process, they result in the generation of relatively large amounts of data per unit of time (Powell et al. 1996). The increased levels of variation detected with these molecular markers has allowed germplasm managers, plant breeders, and geneticists to pursue with new vigor, important questions relating to crops and their genetic resources (Bretting and Widrlechner 1995).

Due to their co-dominant inheritance and amenability to high throughput, SSRs have become a tool of choice for investigations of critical importance to crop germplasm managers, such as the establishment of unique genetic identities or fingerprints, determination of genetic relatedness between accessions, and the assessment of genetic diversity contained within a collection. SSR markers were used by Rongwen et al. (1995) to develop unique DNA profiles or fingerprints for 96 soybean [*Glycine max* (L.) Merr.] cultivars. Seven SSR loci clearly differentiated all but two closely related individuals in the array. Similar results are being reported for other crop species: grape (*Vitis* L.) (Thomas and Scott 1993; Botta et al. 1995; Lamboy and Alpha 1998); apple (*Malus* × *domestica* Borkh.) (Szewc-McFadden et al. 1996; Guilford et al. 1997; sorghum [*Sorghum bicolor* (L.) Moench] (Brown et al. 1996); hops (*Humulus lupulus* L.) (Brady et al. 1996); barley (*Hordeum vulgare* L.) (Russell et al. 1997); avocado (*Persea americana* M.) (Lavi et al. 1994) and cucumbers and melons (*Cucumis sp.*), Katzir et al. 1996). SSRs have also been used to evaluate cultivar variation in rice (*Oryza sativa* L.) (Olufowote et al. 1997).

Genetic relationships have been investigated using SSRs. Provan et al. (1996) found sufficient polymorphism with 16 SSR primers to clearly differentiate 18 potato cultivars (*Solanum tuberosum* L.). However, the use of shared-band analysis to investigate phenetic relationships produced groupings among the cultivars which did not agree with their co-ancestry. The authors concluded that a more appropriate procedure for determining genetic relationships with SSRs would be necessary. In contrast, Liu et al. (1995) studied genetic relationships among 46 ecotypes of *Paspalum vaginatum* Swartz. The relationships revealed were in general agreement with those produced with a previous RAPD analysis. Similarly, Plaschke et al. (1995) used 23 different wheat (*Triticum aestivum* L.) microsatellites to distinguish and estimate genetic diversity among 40 closely related bread wheats. Many of the groups produced in this manner made sense in the light of known pedigree information. In a study of the house mouse (*Mus musculus*), Blouin et al. (1996) demonstrated that 20 unlinked microsatellite loci could readily differentiate unrelated individuals from full sibs and that these two categories could be differentiated from half-sibs greater than 80% of the time. Moreover, in an UPGMA clustering analysis of offspring from four independent half sibships, true family relationships were perfectly represented.

The high levels of variability and reproducibility associated with SSR markers will allow them to serve as anchor markers between different genetic maps within a specific crop (Beckmann and Soller 1990). The SSR markers now being positioned on genetic maps will further facilitate plant breeders' and geneticists' efforts to localize agronomically/horticulturally significant genes or gene complexes to specific sites on the genome (Wu and Tanksley 1993; Akkaya et al. 1995; Akagi et al. 1996; Broun and Tanksley 1996; Liu et al. 1996; Panaud et al. 1996; Senior et al. 1996; Echt and Nelson 1997). From this point, marker-assisted selection or map-based cloning can be utilized to move these traits into desired cultivars (Tanksley and Nelson 1996; Tanksley et al. 1996; Xiao et al., 1996; Tanksley and McCouch 1997).

One of the world's largest *Malus* germplasm collections is curated at the Plant Genetic Resources Unit (PGRU) in Geneva, New York. Over 2500 accessions of apple (*Malus* × *domestica* Borkh.) are maintained as trees in orchards, with an additional 700 accessions of wild *Malus* species stored as seed or maintained as seedlings in the field. Maintaining long-lived woody perennials such as *Malus* is estimated to cost from $50–75 per accession/year, so identifying and eliminating redundancy in such collections is a priority (Forsline and Way 1993).

In order to begin the detailed characterization of *Malus* germplasm resources necessary to identify duplication and/or redundancy in the collection, and to build a database that makes the germplasm more useful to the user community, a core subset collection was established for *Malus* (Kresovich et al. 1995; Forsline 1996). This collection has been planted in a multi-site field replication to allow for evaluations of general and regionally important horticultural traits and biotic and abiotic resistances in several environments (Forsline 1996). As part of this intensified effort to evaluate the *Malus* core subset collection, the present study was initiated to determine genetic identities, estimate genetic diversity, and to identify genetic relationships among *M.* × *domestica* accessions in the core subset using a set of eight SSR primers developed at the PGRU (Szewc-McFadden et al. 1995, 1996).

Table 1 Sequences of *Malus* SSR primer pairs

Locus	5′ Primer sequence	Repeat motif	3′ Primer sequence
GD 12	5′TTGAGGTGTTTCTCCCATTGGA	$(CT)_{32}$	5′CTAACGAAGCCGCCATTTCTTT
GD 15	5′CGAAAGTGAGCAACGAACTCC	$(AGC)_5$	5′ACTCCATCATCGGGTGGTG
GD 96	5′CGGCGGAAAGCAATCACCT	$(TC)_{22}$	5′GCCAGCCCTCTATGGTTCCAGA
GD 100	5′ACAGCAAGGTGTTGGGTAAGAAGGT	$(GA)_{12}$	5′TGCGGACAAAGGAAAAAAAAAAGTG
GD 103	5′CGGCGAGAAAAAAAAACAATG	$(GA)_{20}$	5′GGATAACCGTCCCCCTCTTC
GD 142	5′GGCACCCAAGCCCCTAA	$(TC)_{19}$	5′GGAACCTACGACAGCAAAGTTACA
GD 147	5′TCCCGCCATTTCTCTGC	$(AG)_7$	5′AAACCGCTGCTGCTGAAC
GD 162	5′GAGGCAAGTGACAAAGAAAGATG	$(GA)_{23}$	5′AAAATGTAACAACCCGTCCAAGTG

Materials and methods

Construction of a size-fractionated genomic library

Genomic DNA was isolated from leaves of *M.* × *domestica* cv Golden Delicious using a modified CTAB technique with a subsequent cesium-chloride gradient purification (Sambrook et al. 1989). The resultant genomic DNA was digested with *Taq*I and electrophoresed on an ultra-pure agarose gel (Gibco, BRL). DNA fragments ranging in size from 100 to 500 bp were isolated from an excised gel slice by micropure and microconcentrators (Amicon), ligated to *Cla*I-digested, de-phosphorylated pGEM-7Zf(+) (Promega) and transformed in *Escherichia coli* JM109 cells (Promega) according to the manufacturer's specifications.

Screening the genomic library

Colonies were transferred onto Hybond-N filters (Amersham). The dimeric repeat $(GA)_{10}$; trimeric repeats $(AAC)_8$, $(AAG)_8$, $(TAG)_8$; and tetrameric oligonucleotide repeats $(GATA)_6$, $(ATCA)_6$, $(ATGA)_6$, $(TTAT)_8$ were 3′-end labeled with a digoxigenin-11-ddUTP Genius Chemiluminescence system using the protocol supplied by Boeringer Mannheim. These labeled oligonucleotides were used as probes to screen the Golden Delicious genomic clones (22500 clones with the dimeric probes; 7500 clones with tri- and tetra-meric probes). For the dimeric repeat, hybridization and subsequent washes were conducted under high-stringency conditions (55°C). For the tri- and tetra-meric repeats, the hybridization and wash temperatures were 48°C and 50°C, respectively. Each filter was screened at least twice, and only colonies scored as positive in both screens were sequenced.

Sequencing positive clones

Plasmid DNA containing the di-, tri-, or tetra-meric repeats was isolated using a modified mini-alkaline lysis with Qiagen-tip 20 column purification (Qiagen, Inc.). Clonal DNA was sequenced from M13 priming sites of pGEM 7Zf(+) using the *Taq* DyeDeoxy Terminator Cycle Sequencing Kit on an upgraded model 373 DNA Sequencing System (Applied Biosystems, Inc.).

Primer design

Oligonucleotides complementary to the flanking regions of the repeats were designed by Designer Primer, ver. 1.03, (Research Genetics). Primer design criteria included Tms of 60 to 65°C with no greater than a 3°C difference in Tm between primer pairs. Each pair was designed to produce PCR products ranging from 80 to 300 bp in length (Table 1). Primers were synthesized on a model 392 Nucleic Acids Synthesizer (Applied Biosystems, Inc.). The forward primer of each primer pair was labeled with a fluorescent-dye group: TET (green) – GD 96F and GD 147F; 6-FAM (blue) – GD 12F and GD 100F; and HEX (yellow) – GD 15F, GD 103F, GD 142F, and GD 162F. Use of three dye colors allowed automated detection of fragments arising from multiple loci within one lane simultaneously (multiplexing), as described by Mayrand et al. (1992) (Fig. 1).

SSR Characterization of the *M.* × *domestica* core subset collection

Genomic DNA was extracted from leaves of 66 *M.* × *domestica* accessions (Table 2) using the DNA extraction protocol described by Lamboy and Alpha (1998). PCR amplifications were carried out in a Perkin Elmer 9600 thermocycler. Two multiplexed primer sets, GD 12, 15 and 96 and GD100 and 103, were PCR amplified in 25-μl reaction mixtures containing 25 ng of genomic template DNA, 10–45 pmol of each primer, 0.20 mM of dNTPs, 2.0 mM $MgSO_4$, 1 × ThermoPol reaction buffer (New England BioLabs), and 0.25 units of Deep $Vent_R$ polymerase (New England BioLabs). The multiplexed primer set GD142, 147 and 162 was PCR-amplified in a 20-μl reaction mixture containing 30 ng of genomic template DNA, 20 pmols of each primer pair, 0.20 mM of dNTPs, 2.0 mM $MgSO_4$, 1 × ThermoPol reaction buffer (New England BioLabs), and 0.20 units of Deep $Vent_R$ polymerase (New England BioLabs). Multiplex primer set GD 142, 147 and 162 was amplified using a "touchdown" amplification program (Mellersh and Sampson 1993). Initially, DNA was denatured for 2 min at 94°C followed by two cycles of 94°C for 60 s, 65°C primer annealing for 30 s, and 72°C primer elongation for 45 s. The following 18 cycles had an annealing temperature reduced by 1°C per two cycles. The last five cycles maintained the 55°C annealing temperature. The multiplex primer sets GD 12, GD 15, GD 96 and GD 100 and GD 103 were amplified following a 4-min denaturation at 94°C. Reaction conditions were 25 cycles at 94°C for 60 s, 55°C primer annealing for 120 s, 72°C primer elongation for 120 s and a 10-min extension at 72°C. Amplified products were electrophoresed on an ABI 377 or 373 DNA sequencing system (Applied Biosystems, Inc.) at the USDA-ARS Plant Genetic Resources Conservation Unit's Applied Genetic Analysis Laboratory, Griffin, Ga. or in the Geneva DNA Sequencing Facility, Department of Entomology, Cornell University, Geneva, N.Y., respectively. Each gel run included amplification products from reactions containing: (1) template DNA from the original clone sequenced to design the particular primer pair (positive control), and (2) one lane containing genomic DNA from Golden Delicious, the source for the genomic library construction (size estimator). Automated sizing of the labeled fragments or alleles was determined by mobility units (mu), using the ABI Genescan 672 software ver. 1.2.2–1. relative to an internal lane standard, Genescan 350-Tamara (Applied Biosystems, Inc.), which was loaded with the amplification products in each lane at the beginning of the run.

Data analysis

Typically, the ABI Genescan 672 software assigned non-integer base-pair size values to the detected fragments. These non-integer values varied within a predictable range between gel runs for

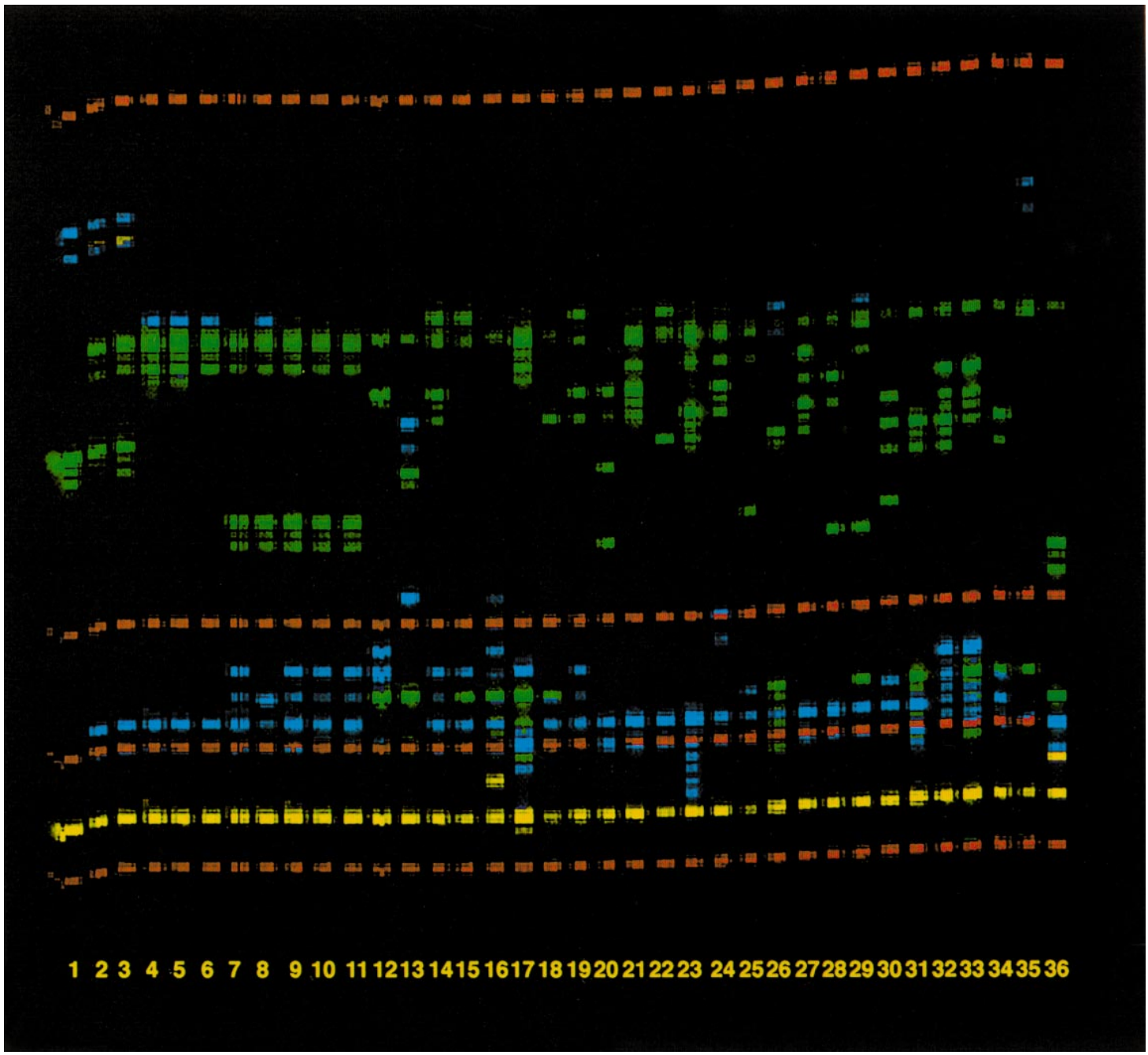


Fig. 1 SSR gel image depicting the reaction products from PCR amplifications of genomic DNA from 32 *Malus* × *domestica* accessions multiplexed with three SSR primer pairs GD 12, (*blue*) 15, (*yellow*) and 96 (*green*). *Red bands* are 350-Tamara internal lane standards sized in base pairs. Going up from the bottom the bands are 139, 150, 160 and 200 base pairs in size, respectively. *Lane 1* contains clonal DNA from the genomic clones out of which the three primer sets were developed. *Lane 2* was loaded with products from the cultivar Golden Delicious and *lane 3* was loaded with products from the indistinguishable accession Chihuahua Gold. *Lanes 4*, *5 and 6* were loaded with products from the indistinguishable McIntosh sports Kimball McIntosh, Marshall McIntosh and Wijcik McIntosh. *Lane 7* was loaded with products from the repository holding of Macfree, while *lane 8* was loaded with products from a Macfree sample obtained from Dan Thompson at the Centre For Plant Health, Sidney, British Columbia, Canada, which was directly traceable to the cultivar's developer. *Lane 9* contains reaction products from the PGRU accession of Nova Easygrow; *lanes 10* and *11* are loaded with products from Nova Easygrow accessions obtained from Dr. Roger Way, Stanley, N.Y., and the Centre for Plant Health, British Columbia respectively. The remainder of the cultivars depicted on the gel are listed in Table 2, with their positions on the gel listed parenthetically

Table 2 Geneva *M.* × *domestica* core subset accessions screened with SSR primers. Parenthetial numbers refer to the lane in which amplification products from these samples were loaded on in Fig. 1

PI	Cultivar	PI	Cultivar
280400	Anna (12)	589894	Keepsake
588995	Antonovka Kamenichke (13)	589122	Kimball McIntosh 2–4-4–4
589956	Antonovka 172670-B	188517	Koningszuur
107196	Antonovka 1.5 pounds	589491	Korichnoe Polosatoje
589726	Britegold	589053	Lady (24)
588835	Burgundy	588943	Liberty (25)
589596	Calville Blanc (14)	589971	Macfree (7)
392311	Chihuahua Gold (3)	588998	Marshall McIntosh (5)
588806	Chisel Jersey	594108	Medaille d'Or
588848	Cortland (15)	588981	Mollie's Delicious (26)
588853	Cox's Orange Pippin	588772	Monroe
589024	Crimson Beauty (16)	589486	Murray
590183	Dayton	588872	Northern Spy (27)
589841	Delicious (17)	588838	Nova Easygrow (9)
589913	Dorsett Golden	590174	Novole (36)
590179	E.8	589970	Petrel
280401	Ein Shemer	383515	Poeltsamaa Winter
123989	Emilia	588798	Rambo-Red Summer
588842	Empire	594111	Redfree
588785	Esopus Spitzenburg (18)	589255	Redspur Delicious
588747	Florina	483257	Reinette Simirenko (28)
588844	Fuji Red Sport Type 2	589520	Rhode Island Greening (29)
392303	Gala (35)	588850	Rome Beauty Law (30)
590184	Golden Delicious (2)	589648	Rosemary Russet (31)
588880	Granny Smith (19)	589845	Smith Jonathan
588837	Gravenstein Wash. Red (20)	589006	Spokane Beauty (32)
589469	Haralson	588955	Sweet Delicious (33)
588841	Idared (21)	589490	Trent
589441	Ingol	589434	Viking
104727	Irish Peach	588778	Virginiagold
246464	James Grieve (22)	590186	Wijcik McIntosh (6)
589962	Jonafree	589645	Winter Majetin
590185	Jonathan (23)	588859	Yellow Transparent (34)

individual genotypes and between genotypes possessing the same fragment. Fragments were assigned integer values which coalesced all the fractional values within a certain size range to one "allele" size or bin, with the computer program 'Binning', written by Lamboy using a Microsoft Fortran Powerstation for IBM-compatible PCs running Windows (Microsoft, Inc., Redmond, Wash.). A complete list of the fragments scored for each genotype in this study can be reviewed on the World Wide Web at the URL http://probe.nalusda.gov:8300/cgibin/browse/rosedb. For this analysis, accessions which showed only one fragment at a locus were considered to be homozygous for that fragment. If in actuality the accession was heterozygous for the fragment and a null allele, the results reported herein underestimate the levels of heterozygosity and gene diversity in the collection.

Allele frequencies, alleles per locus, direct count heterozygosity, polymorphic information content (PIC) (Röder et al. 1995), discrimination power (Jones 1972; Kloosterman et al. 1993), and Nei's genetic identities (Nei 1972) were calculated using the computer program 'SSRS' written by Lamboy using the Microsoft Fortran Powerstation for IBM-compatible PCs running Windows. Effective alleles per locus (A_{ep}) were calculated according to Weir (1989) with the formula $1/(1 - H_{ep})$, where H_{ep}, the genetic diversity per locus, is equal to $1 - \sum p_i^2$ and p_i is equal to the frequency of the i^{th} allele at the locus. Direct count heterozygosities were calculated as the number of genotypes which were heterozygous at a given locus divided by the total number of genotypes scored at that locus. PIC was calculated with the following formula, $1 - \sum_{i=1}^{n} p_i^2$, where p_i equals the frequency of the i^{th} allele. The discrimination power at a locus, which provides an estimate of the probability that two randomly sampled accessions in the study would be differentiated by their allelic profiles, was obtained for both the sample under investigation and an infinitely large theoretical population with the same genotype frequencies found in the sample population. The value was calculated with the formula $1 - \Sigma(P_i)^2$, where P_i represents the frequency of each genotype (Kloosterman et al. 1993).

Genetic relationships among the 66 accessions in this study were investigated using an unweighted pair-group method (UPGMA) cluster analysis of Nei's genetic identities for the accessions (Sneath and Sokal 1973). The analysis and a phenogram (see Fig. 3) were computed with the program NTSYS-pc, ver. 1.80 (Rohlf 1994).

Results

Frequency of di-, tri-, and tetra-nucleotide repeats in the apple genome

The average insert size for the *M.* × *domestica* Golden Delicious genomic library was approximately 300 bp in length. Among the 22 500 clones screened with the dimeric probe and the 7500 clones screened with tri and tetra-meric probes, 103, 4, and 1 positive clones were detected, respectively. Using these numbers, we calculate that a $(GA)_{10}$ repeat occurs every 87 kb; $(AAC)_8$, $(AAG)_8$, or $(TAG)_8$ repeats occur every 562.5 kb; and $(GATA)_6$, $(ATCA)_6$, $(ATGA)_6$, or $(TTAT)_8$ repeats occur every 2250 kb. Based on an estimated 2n genome size for *M.* × *domestica* of 1.55 ρg DNA/2C (Dickson et al. 1992), converted to 3.1×10^6 kb, we estimate

Table 3 SSR primer product characterization

Locus	Expected product size	Range of product sizes	Number putative alleles	A_{ep}^{a}	Direct count heterozygosity	PIC
GD 12	192 bp	141–191 bp	12	3.95	0.758	0.747
GD 15	144 bp	144–147 bp	2	1.02	0.015	0.015
GD 96	173 bp	152–197 bp	15	7.51	0.909	0.867
GD 100	227 bp	223–242 bp	14	7.27	0.879	0.862
GD 103	108 bp	90–133 bp	13	3.73	0.333	0.732
GD 142	143 bp	123–158 bp	13	8.55	0.909	0.883
GD 147	138 bp	124–156 bp	15	4.67	0.848	0.786
GD 162	234 bp	215–254 bp	13	5.07	0.894	0.803
		$\bar{\times}$	12.13	5.22	0.693	0.712

[a] Effective alleles per locus

the total number of our di-, tri-, and tetra-meric repeats in the apple genome to be 3.6×10^4, 8.6×10^3, and 1.4×10^3, respectively.

Genetic diversity

All eight primer pairs generated multiple fragments in this group of *Malus* accessions. DNA from three apple mapping populations has also been amplified with the eight primer pairs, and many of the fragments have been observed to segregate in a Mendelian fashion (Minou Hemmat, personal communication). Henceforth the primer pairs will be referred to as loci and the fragments generated by each primer pair in an amplification reaction will be called alleles. The number of alleles detected per locus ranged from two for GD 15 to 15 for both GD 96 and GD 147, with a mean value over all eight loci of 12.1 alleles per locus (Table 3). A total of eight homozygous null alleles in the 528 possible accession by loci combinations were detected in this study. Generally, frequencies for individual alleles were low, with one allele occurring at a higher frequency at each locus (Fig. 2).

Direct-count heterozygosities for individual loci ranged from 0.015 for GD 15 to 0.909 for GD 96 and GD 142, with the average direct count heterozygosity for all loci in the study equalling 0.693 (Table 3). Genetic diversity or polymorphic information content (PIC) per locus ranged from 0.015 at GD 15 to 0.883 for GD 142, with an average PIC value for all loci of 0.712 (Table 3). Generally, PIC values increased proportionally with increasing heterozygosity at a locus. However, because rare alleles have less influence on PIC values than common alleles, this trend was not consistent. For example, although loci GD 96 and GD 142 both had direct count heterozygosities of 0.909, the PIC values were different (0.867 vs 0.883 respectively). At locus GD 96 the average occurrence among the 66 genotypes in this study for the 15 fragments detected was 8.7, whereas for GD 142 the average occurrence for the 13 fragments detected in the population was 10.0, which accounted for the higher PIC score at GD 142 (Table 3; Fig. 2).

Genetic identities

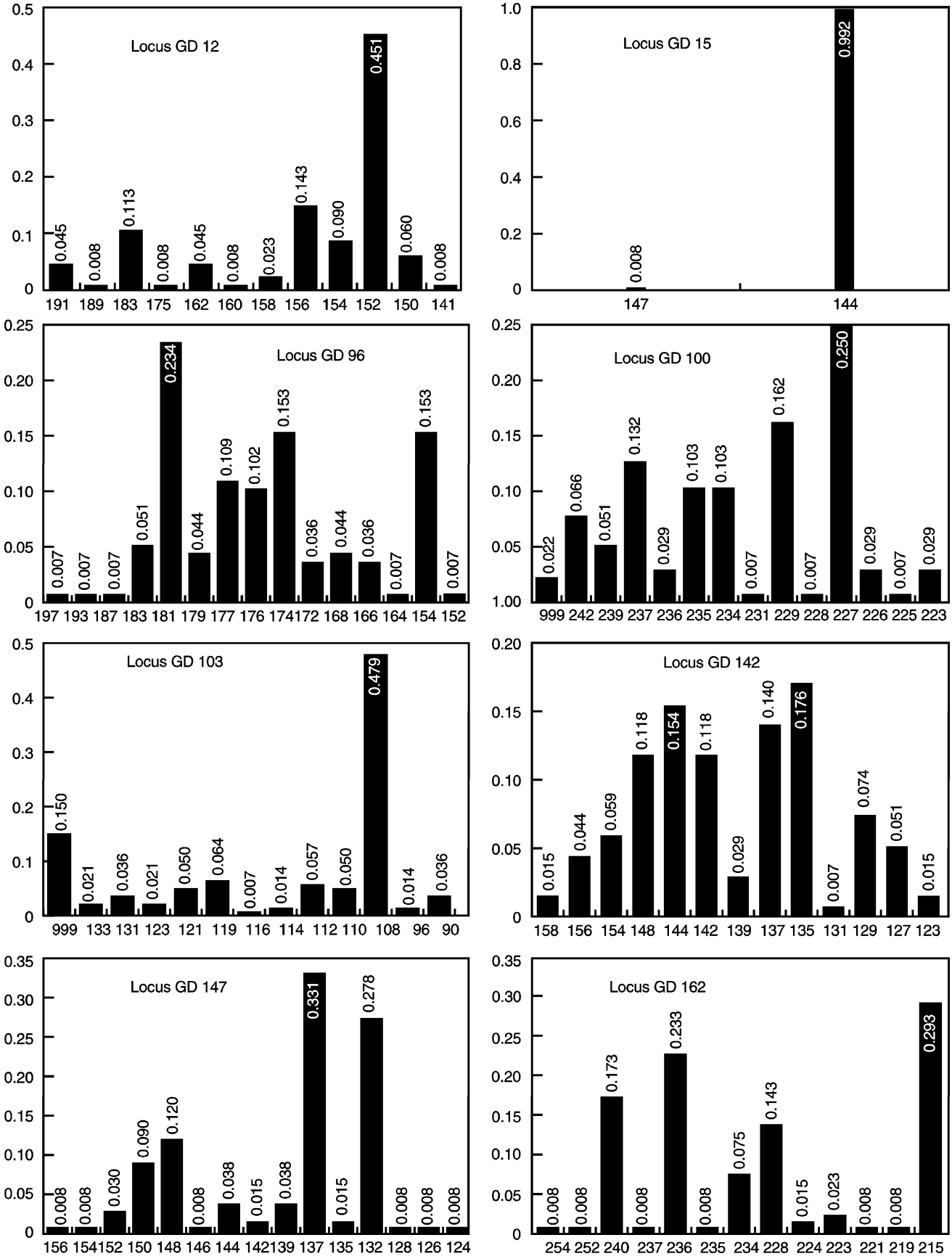
The eight primer pairs utilized in this study unambiguously discriminated all but seven pairs of accessions in this collection of 66 *M.* × *domestica* genotypes. In fact, 1381 of the 2145 possible pairwise comparisons for the 66 accessions differed at seven or eight of the loci used in this study. Of the seven indistinguishable genotype pairs, five were sport mutations and/or their progenitors: namely, Marshall McIntosh and Kimball McIntosh; Marshall McIntosh and Wijcik McIntosh; Kimball McIntosh and Wijcik McIntosh; Redspur Delicious and Delicious; and Jonathan and Smith Jonathan. The remaining two indistinguishable pairs were Nova Easy Gro and Macfree; and Chihuahua Gold and Golden Delicious.

The high discrimination power of the loci (Table 4) suggests that these eight primer pairs could differentiate any two accessions within the *M.* × *domestica* collection. The discrimination power at individual loci in the sample ranged from 0.030 for GD 15 to 0.974 for GD 96 (Table 4). The probability of matching any two genotypes at a locus in the sample ranged from 0.026 at GD 96 to 0.970 at GD 15. The probability of matching any two genotypes at all loci by chance in the sample was 0.156×10^{-8} (Table 4).

Genetic relationships

The majority rule consensus phenogram generated from the UPGMA cluster analysis of the 66 accessions

Fig. 2 Eight histograms depicting the alleles which occurred at each locus in this collection of 66 *M.* × *domestica* accessions and the frequency at which each allele occurred


Locus GD 12
0.045 0.008 0.113 0.008 0.045 0.008 0.023 0.143 0.090 0.451 0.060 0.008
191 189 183 175 162 160 158 156 154 152 150 141
Locus GD 15
0.008 0.992
147 144
Locus GD 96
0.007 0.007 0.007 0.051 0.234 0.044 0.109 0.102 0.153 0.036 0.044 0.036 0.007 0.153 0.007
197 193 187 183 181 179 177 176 174172 168 166 164 154 152
Locus GD 100
0.022 0.066 0.051 0.132 0.029 0.103 0.103 0.007 0.162 0.007 0.250 0.029 0.007 0.029
999 242 239 237 236 235 234 231 229 228 227 226 225 223
Locus GD 103
0.150 0.021 0.036 0.021 0.050 0.064 0.007 0.014 0.057 0.050 0.479 0.014 0.036
999 133 131 123 121 119 116 114 112 110 108 96 90
Locus GD 142
0.015 0.044 0.059 0.118 0.154 0.118 0.029 0.140 0.176 0.007 0.074 0.051 0.015
158 156 154 148 144 142 139 137 135 131 129 127 123
Locus GD 147
0.008 0.008 0.030 0.090 0.120 0.008 0.038 0.015 0.038 0.331 0.015 0.278 0.008 0.008 0.008
156 154152 150 148 146 144 142139 137 135 132 128 126 124
Locus GD 162
0.008 0.008 0.173 0.008 0.233 0.008 0.075 0.143 0.015 0.023 0.008 0.008 0.293
254 252 240 237 236 235 234 228 224 223 221 219 215

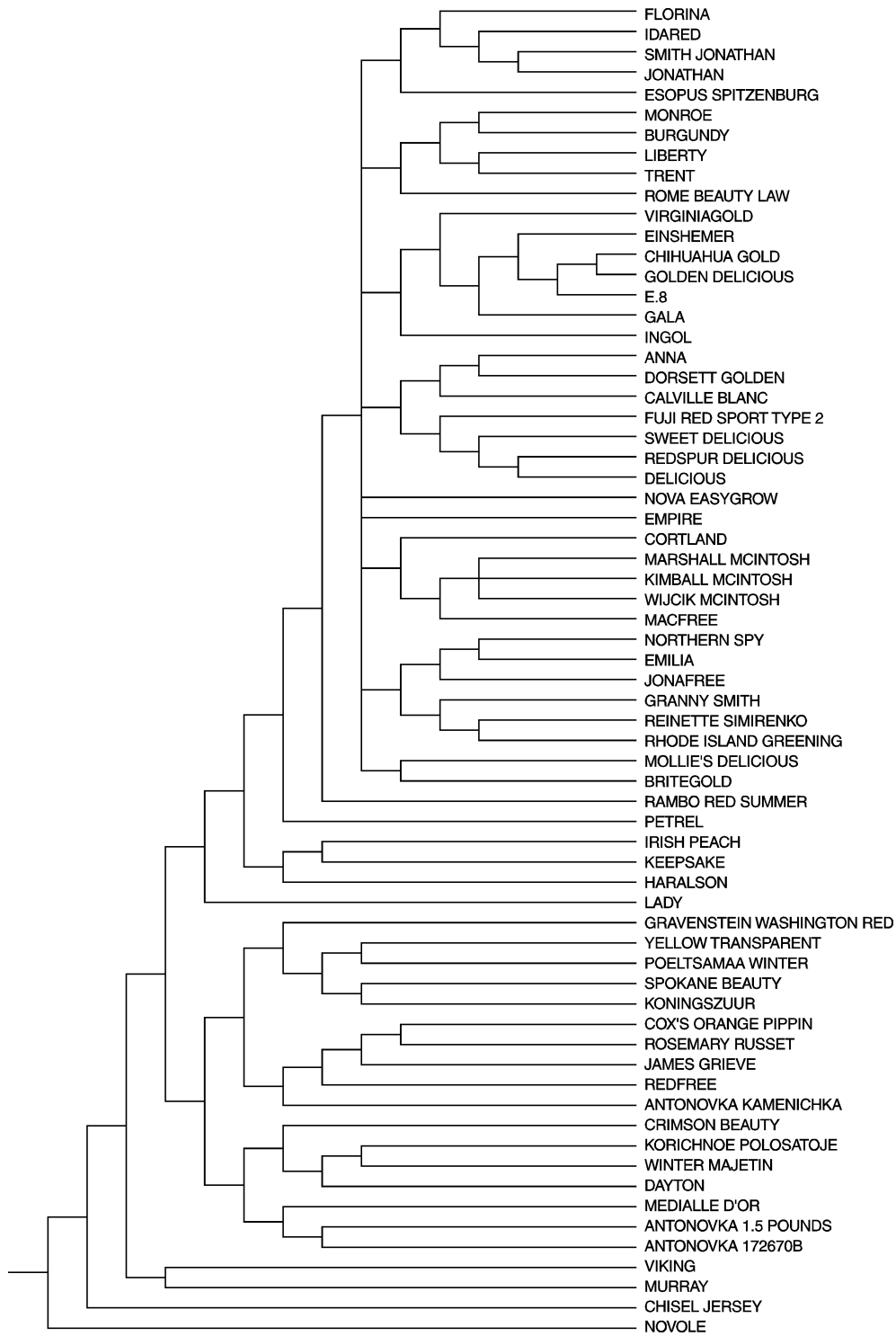


Fig. 3. Majority rule consensus phenogram for the 66 *M.* × *domestica* accessions evaluated in this study. The phenogram was produced using an UPGMA cluster analysis of Nei's genetic identities between the accessions

in this study resulted in several groupings which appear to be related to their pedigrees (Fig. 3). The first group, which includes Florina through Trent, are all cultivars which have at least one-quarter Jonathan in their pedigrees, with the exception of Liberty, which has no Jonathan in its pedigree. This group also contains the old cultivar Esopus Spitzenburg, which is generally considered to be the maternal parent of Jonathan. The second grouping, which includes Virginia Gold through Dorsett Golden, all had

Table 4 SSR genotype differentiation capacity within the *Malus* × *domestica* core collection

Locus	Discrimination power	Probability of matching genotype
GD 12	0.918	0.825×10^{-1}
GD 15	0.030	0.970
GD 96	0.974	0.256×10^{-1}
GD 100	0.967	0.331×10^{-1}
GD 103	0.838	0.162
GD 142	0.972	0.280×10^{-1}
GD 147	0.935	0.648×10^{-1}
GD 162	0.922	0.779×10^{-1}
For all loci	1.00	0.156×10^{-8}

a minimum of one-half Golden Delicious in their ancestry. The next group, Fuji Red Sport #2 through Delicious, are all Delicious types containing a minimum of one-half Delicious in their pedigrees. The cultivars Nova Easygrow through Mcfree constituted a group with at least one-quarter McIntosh in their ancestry. Two cultivars, Emilia and Jonafree, grouped around the cultivar Northern Spy. Emilia is one-half Northern Spy while Jonafree contains one-quarter Red Spy, a sport mutation of Northern Spy, in its pedigree. The final large group in the phenogram, Granny Smith through Chisel Jersey, contains a number of old Russian and European cultivars, most of which are chance discoveries of unknown parentage.

Discussion

We estimate that the $(GA)_{10}$ repeat occurs every 87 kb in the apple genome, which falls within the values currently reported in the literature for a range of taxa. These intervals range from 60 kb to every 1.2 Mb (Condit and Hubbell 1991; Lagercrantz et al. 1993; Wu and Tanksley 1993; Röder et al. 1995; Broun and Tanksley 1996; Chase et al. 1996; Liu et al. 1996; Panaud et al. 1996; Guilford et al., 1997). In their study of 21 apple cultivars, Guilford et al. (1997) estimate that the $(GA)_{15}$ repeat occurs every 120 kb in apple. The differences in our estimates could be accounted for by their evaluating a much larger portion of the genome. Based on a genome size estimate for apple of 1.5×10^6 kb, with an average insert size for the 5000 genomic clones they screened of 15 kb, they looked at approximately 0.05% of the apple genome compared to the 0.0045% we evaluated. Lavi et al. (1994) reported a high frequency of dinucleotide repeats in avocado; one AG every 7.9 kb, one CA every 20 kb, and a GC every 26 kb. In the avocado study only 800 kb of the genome, 0.0004% of the total, was screened, which lends further support to the notion that an overestimate of the frequency of the repeats can result from an undersampling of the genome.

However, the fact that we used smaller insert clones in the library allowed us to detect a number of repeats which may have been effectively masked within the larger insert clones used by Guilford et al. (1997). The same kind of relationship can be seen in our respective estimates for the occurrence of trimeric repeats in the apple genome. Our estimate of every 562.5 kb is much higher than the estimate of 3 Mb they make. Similar inverse relationships between the size of the clones used in the genomic library and the frequency of occurrence of the repeats detected were noted by Condit and Hubbell (1991) and Liu et al. (1995).

The mean number of alleles detected per locus in our study ranged from 2 to 15, with a mean value of 12.1 alleles per locus. This value is high when compared to the plant SSR literature at large. However, many of the low-end values reported in the plant literature are for self-pollinating and/or annual crops such as tomato, with estimates ranging from 1.5 to 3.1 mean alleles per locus (Broun and Tanksley 1996; Smulders et al. 1997 respectively); wheat, with estimates of 3.8, 4.6 and 6.2 (Devos et al. 1995; Röder et al. 1995; Plaschke et al. 1995 respectively); sorghum 2.3 (Brown et al. 1996); cucumbers and melons 2.6 and 2.9 (Katzir et al. 1996) and watermelons with 2.0 alleles per locus (Jarret et al. 1997). Our estimates are similar to values reported for plant species with similar life-history characteristics, including outcrossed, long-lived woody perennials such as grape, with reported values of 8.4 and 27.6 alleles per locus (Thomas and Scott 1993; Lamboy and Alpha 1998); avocado, 9.5 (Lavi et al. 1994); tropical trees, 6.4 (Chase et al. 1996); *Pinaceae*, 6.0, 5.4, 8.2 and 13.0 (Smith and Devey 1994; Echt et al. 1996; van de Ven and McNicol 1996; Pfeiffer et al. 1997); oak, 14.3 (Dow et al. 1995); and citrus, 5.5 (Kijas et al. 1995).

Our alleles per locus estimate is considerably higher than the 4.5 alleles per locus reported by Guilford et al. (1997) in their study of 21 apple cultivars. Several factors could account for the discrepancy in our estimates, the most likely may simply be the differences in our sample sizes. We looked at approximately three times the number of genotypes (66), thereby increasing the likelihood that we would encounter more allelic diversity. Moreover, the core subset we evaluated was designed to maximize allelic diversity. In addition, we screened our primers at the outset of the project to find primer pairs which detected maximum amounts of variability. In the Guilford study, two primers were monomorphic and half of their primers (5) revealed four or fewer alleles in the sample population. Finally, it is possible that the automated detection system utilized in the present study was capable of resolving allelic variation at a finer scale than the probe hybridization detection system used in the Guilford study.

The direct-count heterozygosities for individual loci ranged from 0.015 to 0.909 for the eight loci in this

study with a mean value for all loci of 0.693 (Table 3). This value falls within the range of values being reported for plant SSR studies and is quite similar to the 0.66 value reported by Guilford et al. (1997) in their study of 21 apple cultivars.

The present study reinforces the utility of SSR primers for providing unique genetic identities or fingerprints of individual plant cultivars or genotypes. The combined discrimination power for all eight loci was effectively 1.0, with the probability of matching any two genotypes at all eight loci in this study being approximately 1 in 1 billion! Of the seven pairs of genotypes we were not able to differentiate, five were sport mutations and/or their progenitors. If these sports arose as the result of one or a small number of mutations, it is not surprising that these small genomic differences were not detectable with our markers. The remaining two indistinguishable pairs, Nova Easygrow and Macfree and Chihuahua Gold and Golden Delicious, presented more complex problems.

According to pedigree information Nova Easygrow and Macfree should have been easily differentiated with our eight SSR markers. Several amplification re-runs with the original DNA extractions and with newly extracted DNA from the same genotypes produced the same results; products from the amplifications were indistinguishable. In addition, no distinguishing characters were found between the two accessions based on leaf and flower phenology and morphology. New budwood of each accession was obtained from independent sources. DNA was extracted from expanding leaves from each and amplification products from these reactions were compared with the original indistinguishable products from the initial reactions (Fig. 1). The Macfree accession obtained from the Centre For Plant Health, Sidney, B.C., Canada, which could be traced directly to the Smithfield Experimental Farm where it was developed, was different from all the other accessions (Fig. 1, lanes 7–11) and was determined to be the authentic Macfree. All three Nova Easygrow accessions used in this experiment proved to be identical and these three were identical to the repository Macfree accession which had been obtained from Rutgers University, Piscataway, N.J. Thus, it appears that our Macfree accession was mislabled. We are currently in the process of re-propagating the authentic genotype.

Chihuahua Gold, a cultivar acquired from Chihuahua, Mexico, came into the repository collection with very little passport information. Reportedly discovered in Guerrero, Chihuahua, Mexico, it has no published pedigree. Lacking an independent sample directly traceable to the original source, it was not possible to check the authenticity of this accession. The information available does not preclude the possibility that Chihuahua Gold is in actuality Golden Delicious rediscovered, or else a sport mutation of Golden Delicious.

The extremely low probability of randomly matching genotypes in this study and a similar study on a collection of *Malus* species and hybrids (Hokanson et al., in preparation) has given us great confidence in the ability of SSR markers to detect misidentified accessions. The markers are so robust in this respect that when accessions with matching genotypes are discovered in the collection, they merit an immediate closer examination. In this manner we have discovered the misidentified accession in this study, along with two others in our species and hybrids collection (Hokanson et al. 1997 b). Likewise, in a study of *Vitis* cultivars and species curated at the PGRU, (Lamboy and Alpha, 1998), a misidentified grape species was discovered using SSRs.

The UPGMA cluster analysis of the 66 accessions in the present study produced several meaningful groupings based on pedigree and/or geographical origins of the accessions. While the positioning of some accessions in the phenogram raises new questions, some of the clustering answers outstanding questions concerning the origins and pedigrees of some old cultivars. The first grouping containing Jonathan types has two interesting members. Although Liberty consistently falls within this group, it's pedigree contains no Jonathan. The fact that Liberty is one-quarter McIntosh suggests that it should align more closely with the McIntosh group. A closer examination of the pedigrees of some closely aligned members of the Jonathan group helps explain this apparent contradiction. The cultivar Trent contains Rome Beauty and *Malus floribunda* 821 in nearly the same proportion as does Liberty. Both of these cultivars align consistently with Rome Beauty Law in the phenogram. The cultivar Burgundy was developed in the same time period in the Geneva, N.Y., breeding program as Liberty. Both Burgundy and Liberty have Rome Beauty and Macoun in their pedigrees. Conversely, the inclusion of Rome Beauty Law in this Jonathan group would be explained by the presence of Rome Beauty in the pedigrees of all the aforementioned cultivars. Interestingly, the consistent positioning of the old cultivar Esopus Spitzenburg in this Jonathan group further solidifies the long-standing claim that Esopus Spitzenburg is the maternal parent of Jonathan. An unanswered question raised by the analysis concerns the old cultivar Calville Blanc, which was consistently located between the Golden Delicious and Red Delicious groups. Despite this consistent alignment, nothing in the pedigrees or the geographic origins of the closely positioned cultivers, Dorsett Golden, Anna, or the Delicious types explains this arrangement.

The final large group of old Russian and European cultivars consists mainly of individuals of unknown parentage. This group, comprising cultivars Granny Smith through Novole, clusters distinctly from the more domesticated types and within it there exist several interesting pairings and a couple of inexplicable alignments. One consistent group is a pair of old

Minnesota cultivars, Haralson and Keepsake, which are one-half and one-quarter Melinda respectively. Interestingly, the breeding program in Minnesota was begun by collecting open-pollinated seed from regionally adapted cultivars growing in local orchards, including Melinda and some old Russian varieties (Jim Luby, personal communication). The close alignment of Haralson and Keepsake within this group of old European and Russian varieties could be explained by the origin of the Minnesota apple-breeding program. Another intriguing story involves the consistent close affiliation of the cultivars Cox's Orange Pippin and James Grieve. The latter is believed to be a seedling of either Cox's Orange Pippin or Potts Seedling. Evaluation of the polymorphisms between the two purported parents of James Grieve would probably help elucidate the true origin of this significant variety in the same manner used by Russell et al. (1997) in determining the true parentage of the barley variety Golden Promise.

The European/Russian cluster also contains several cultivars whose presence is not explained by their pedigrees. The cultivar Murray is a seedling of McIntosh and yet it shows no affiliation with that group. The cultivars Redfree and Dayton are both one-eighth Jonathan with some McIntosh in their backgrounds, but they do not align with either of those groups. Novole, an apple rootstock cultivar, was the consistent outlier in the analysis. It is thought to be a hybrid between the apple species *Malus prunifolia* and *Malus sieboldii*, which would account for its disparate genetic relationship with the other apple cultivars in this study.

Great expectations have accompanied the early use of simple sequence repeat (SSR) markers in plant genetics research. Our initial look at *Malus* germplasm in the USDA National Plant Germplasm Systems collection with eight SSRs developed at the Plant Genetic Resources Unit has demonstrated that the expectations were not unrealistic for apple. The markers have allowed us to make estimates of overall genetic diversity within the collection, assign unique genetic fingerprints to nearly all accessions in the *M. domestica* core while simultaneously revealing molecular-based genetic relationships which are meaningful in terms of known pedigree information. We have also been able to uncover previously unidentifed, mislabeled accessions.

Results from this study and a similar study of our species and hybrid collection (Hokanson et al., manuscript in preparation) will help to determine curatorial decisions regarding issues such as de-accessioning in the ex situ *Malus* germplasm collection. In addition, these results will be compared to those being generated from a similar survey of wild *Malus sieversii* (Ledeb.) M. Roemer, collected from the center of diversity in Kazakstan (Hokanson et al. 1997 a). Such a comparison will also direct future collection efforts in Central Asia and guide endeavors to establish in situ germplasm reserves for *M. sieversii* which preserve maximal amounts of genetic diversity for the species.

Acknowledgments We thank Joseph Ficcaglia, Dana Sauer and Chris Alpha for providing excellent technical assistance and Drs. Steve Kresovich and Doug Knipple for allowing us to use their automated gel-electrophoresis equipment. In particular we thank Dr. Steve Kresovich for his advice, assistance, and encouragement, without which this project would not have come to fruition. In addition, we thank Dan Thompson at the Centre for Plant Health, Sidney, British Columbia, Canada, for providing plant material, Drs. Susan Brown and Roger Way for providing pedigree information and reviewing portions of the manuscript, Dr. Norman Weeden for providing segregating populations and mapping data, and Drs. Perry Cregan, James Hancock and Stephan Krebs for excellent critical reviews of this manuscript. Mention of a proprietary product in this manuscript does not constitute a product endorsement by the authors or the USDA, nor implied criticism of equally suitable products not mentioned herein.

References

Akagi H, Yokozeki Y, Inagaki A, Fujimura T (1996) Microsatellite DNA markers for rice chromosomes. Theor Appl Genet 93 : 1071–1077.

Akkaya MS, Shoemaker RC, Specht JE, Bhagwat AA, Cregan PB (1995) Integration of simple sequence repeat DNA markers into a soybean linkage map. Crop Sci 35 : 1439–1445

Beckmann JS, Soller, M (1990) Toward a unified approach to genetic mapping eukaryotes based on sequence tagged microsatellite sites. Bio/Technology 8 : 930–932

Blouin MS, Parsons M, Lacaille V, Lotz S (1996) Use of microsatellite loci to classify individuals by relatedness. Mol Ecol 5 : 393–401

Botta L, Scott NS, Eynard I, Thomas MR (1995) Evaluation of microsatellite sequence tagged site markers for characterizing *Vitis vinifera* cultivars. Vitis 34 : 99–102

Brady JL, Scott NS, Thomas MR (1996) DNA typing of hops (*Humulus lupulus*) through application of RAPD and microsatellite marker sequences converted to sequence tagged sites (STS). Euphytica 91 : 277–284

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

Broun P, Tanksley SD. (1996) Characterization and genetic mapping of simple repeat sequences in the tomato genome. Mol Gen Genet 250 : 39–49

Brown AHD (1989 a) 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, UK, pp 136–156

Brown AHD (1995) The core collection at the crossroads. In: Hodgkin T, Brown AHD, Hintum TJL, Morales, EAV (eds) Core Collections of Plant Genetic Resources. John Wiley and Sons, Chichester, UK, pp 3–19

Brown SM, Hopkins MS, Mitchell SE, Senior ML, Wang TY, Duncan RR, Gonalez-Candelas F, Kresovich S (1996) Multiple methods for the identification of polymorphic simple sequence repeats (SSRs) in sorghum [*Sorghum bicolor* (L.) Moench]. Theor Appl Genet 93 : 190–198

Brush SB (1989) Rethinking crop genetic resource conservation. Biol Conserv 3 : 1–11

Chapman C (1989) Principles of germplasm evaluation. In: Stalker HT, Chapman C. (eds) IBPGR Training Courses: Lecture Series 2. Scientific Management of germplasm: Characterization, Evaluation, and enhancement. International Board for Plant Genetic Resources, Rome, pp 55–64.

Chase M, Kesseli R, Bawa K (1996) Microsatellite markers for population and conservation genetics of tropical trees. Am Jour Bot 83 : 51–57

Clegg MT (1990) Molecular diversity in plant populations. In: Brown AHD, Clegg MT, Kahler AL, Weir BS (eds) Plant population

genetics, breeding, and genetic resources. Sinauer Associates Inc., Sunderland, Massachusetts, pp 98–115
Condit R, Hubbell JP (1991) Abundance and DNA sequence of two-base repeat regions in tropical tree genomes. Genome 34 : 66–71
Devos KM, Bryan GJ, Collins AJ, Stephenson P, Gale MD. (1995) Application of two microsatellite sequences in wheat storage proteins as molecular markers. Theor Appl Genet 90 : 247–252
Dickson EE, Arumuganathan K, Kresovich S, Doyle JJ (1992) Nuclear DNA content variation within the Rosaceae. Am Jour Bot 79 : 1081–1086
Doebley J (1989) Isozymic evidence and the evolution of crop plants. In: Soltis DE, Soltis PS (eds) Isozymes in plant biology. Dioscorides Press, Portland, Oregon, pp 165–191
Dow BD, Ashley MV, Howe HF (1995) Characterization of highly variable $(GA/CT)_n$ microsatellites in the bur oak (*Quercus macrocarpa*). Theor Appl Genet 91 : 137–141
Echt CS, Nelson CD (1997) Linkage mapping and genome length in eastern white pine (*Pinus strobus* L.). Theor Appl Genet 94 : 1031–1037
Echt CS, May-Marquardt P, Hseih M, Zahorchak R. (1996) Characterization of microsatellite markers in eastern white pine. Genome 39 : 1102–1108
Forsline PL (1996) Core subsets in the USDA/NPGS with Apple as an example. In: Davidson CG, Warner J (eds) Proc 2nd Wkshp Clonal Genetic Resources: emerging issues and new directions. pp 172–175
Forsline PL, Way RD (1993) Apple accessions of low priority targeted for removal from the National Plant Germplasm System. Fruit Var Jour 47 : 204–214
Frankel OH (1984) Genetic perspectives of germplasm conservation. In: Arber W, Llimensee K, Peacock WJ, Starlinger, P (eds) Genetic Manipulation: impact on man and society. Cambridge, UK: Cambridge University Press, pp
Gepts P (1990) Genetic diversity of seed storage proteins in plants. In: Brown AHD, Clegg MT, Kahler AL, Weir BS (eds) Plant population genetics, breeding and genetic resources. Sinauer Associates Inc., Sunderland, Massachusetts, pp 64–82
Gepts P (1995) Genetic markers and core collections. In: Hodgkin T, Brown AHD, Hintum ThJL, Morales EAV (eds) Core collections of plant genetic resources. John Wiley and Sons, Chichester, UK, pp 127–146
Guilford P, Prakash S, Zhu JM, Rikkerink E, Gardiner S, Bassett H, Forster R (1997) Microsatellites in *Malus* × *domestica* (apple): abundance, polymorphism and cultivar identification. Theor Appl Genet 94 : 249–254
Hilu KW (1989) Taxonomy of cultivated plants. In: Stalker HT, Chapman C (eds) IBPGR training courses: lecture series 2. Scientific management of germplasm characterization, evaluation and enhancement. International Board for Plant Genetic Resources, Rome, pp 33–40
Hokanson SC, McFerson JR, Forsline PL, Lamboy WF, Luby HS, Aldwinckle HS, Djangaliev AD (1997 a) Collecting and managing wild *Malus* germplasm in its center of diversity. HortScience 32 : 173–176
Hokanson SC, Szewc-McFadden AK, Lamboy WF, McFerson JR (1997 b) Simple sequence repeat (SSR) variation in a collection of *Malus* species and hybrids. HortScience 32 : 440
Jarret RL, Merrick LC, Holms T, Evans J, Aradhya MK (1997) Simple sequence repeats in watermelon [*Citrullus lanatus* (Thunb.) Matsum. & Nakai]. Genome 40 : 433–441
Jones DA (1972) Blood samples: probability of discrimination. J Forens Sci Soc 12 : 355–359
Katzir N, Danin-Poleg Y, Tzuri G, Karchi Z, Lavi U, Cregan PB. (1996) Length polymorphism and homologies of microsatellites in several *Cucurbitaceae* species. Theor Appl Genet 93 : 1282–1290
Kijas JMH, Fowler JCS, Thomas MR (1995) An evaluation of sequence-tagged microsatellite site markers for genetic analysis within *Citrus* and related species. Genome 38 : 349–355
Kloosterman AD, Budowle B, Daselaar P (1993) PCR-amplification and detection of the human D1S80 VNTR locus. Int J Leg Med 105 : 257–264
Kresovich S, McFerson, JR (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
Kresovich S, Lamboy WF, McFerson JR, Forsline PL (1995) Integrating different types of information to develop core collections, with particular reference to *Brassica oleracea* and *Malus* × *domestica*. In: Hodgkin T, Brown AHD, Hintum TJL, Morales EAV (eds) Core collections of plant genetic resources. John Wiley and Sons, Chichester, UK, pp 147–154
Lagercrantz U, Ellegran H, Andersson L (1993) The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. Nucleic Acids Res 21 : 1111–1115
Lamboy WF, Alpha CG (1988) The utility of simple sequence repeats (SSRs) for DNA fingerprinting germplasm accessions of grape (*Vitis* L.) species. J Am Soc Hortic Sci 123 : 182–188
Lavi U, Akkaya M, Bhagwat A, Lahav E, Cregan PB (1994) Methodology of generation and characteristics of simple sequence repeat DNA markers in avocado (*Persea americana* M.). Euphytica 80 : 171–177
Liu ZW, Jarret RL, Kresovich S, Duncan RR (1995) Characterization and analysis of simple sequence repeat (SSR) loci in seashore paspalum (*Paspalum vaginatum* Swartz.). Theor Appl Genet 91 : 47–52
Liu ZW, Biyashev RM, Saghai Maroof MA (1996) Development of simple sequence repeat DNA markers and their integration into a barley linkage map. Theor Appl Genet 93 : 869–876
Lyman JK (1984) Progress and planning for germplasm conservation of major food crops. Plant Genet Res Newslett 60 : 3–21
Marshall DR (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. John Wiley and Sons, Chichester, UK, pp 367–388
Mayrand EP, Corcoran KP, Ziegle JS, Robertson JM, Hoff LB, Kronick MN (1992) The use of fluorescence detection and internal lane standards to size PCR products automatically. Appl Theor Electrophoresis 3 : 1–11
Mellersh C, Sampson J (1993) Simplifying detection of microsatellite length polymorphisms. BioTechniques 15 : 582–584.
Nabhan GP (1985) Native crop diversity in Aridoamerica: conservation of regional gene pools. Econ Bot 39 : 387–399
Nei M (1972) Genetic distance between populations. Am Nat 106 : 283–292
Olufowote JO, Xu Y, Chen X, Park WD, Beachell HM, Dilday RH, Goto M, McCouch SR (1997) Comparative evaluation of within-cultivar variation of rice (*Oryza sativa* L.) using microsatellite and RFLP markers. Genome 40 : 370–378
Panaud O, Chen X, McCouch SR (1996) Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (*Oryza sativa* L.). Mol Gen Genet 252 : 597–607
Pfeiffer A, Olivieri AM, Morgante M (1997) Identification and characterization of microsatellites in Norway spruce (*Picea abies* K.). Genome 40 : 411–419
Plaschke J, Ganal MW, Röder MS (1995) Detection of genetic diversity in closely related bread wheat using microsatellite markers. Theor Appl Genet 91 : 1001–1007
Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol Breed 2 : 225–238
Provan J, Powell W, Waugh R (1996) Microsatellite analysis of relationships within cultivated potato (*Solanum tuberosum*). Theor Appl Genet 92 : 1078–1084
Röder MS, Plaschke J, König SU, Börner A, Sorrells ME, Tanksley SD, Ganal MW (1995) Abundance, variability and

chromosomal location of microsatellites in wheat. Mol Gen Genet 246 : 327–333
Rohlf FJ (1994) NTSYS-pc, version 1.80. Exeter Software, Setauket, New York
Rongwen J, Akkaya MS, Bhagwat AA, Lavi U, Cregan PB (1995) The use of microsatellite DNA markers for soybean genotype identification. Theor Appl Genet 90 : 43–48
Russell J, Fuller J, Young G, Thomas B, Taramino G, Macaulay M, Waugh R, Powell W (1997) Discriminating between barley genotypes using microsatellite markers. Genome 40 : 442–450
Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
Senior ML, Chin ECL, Lee M, Smith JSC, Stuber CW (1996) Simple sequence repeat markers developed from maize sequences found in the GENBANK database: Map construction. Crop Sci 36 : 1676–1683
Smith DN, Devey ME (1994) Occurrence and inheritance of microsatellites in *Pinus radiata*. Genome 37 : 977–983
Smulders MJM, Bredemeijer G, Rus-Kortekaas W, Arens P, Vosman B (1997) Use of short microsatellites from database sequences to generate polymorphisms among *Lycopersicon esculentum* cultivars and accessions of other *Lycopersicon* species. Theor Appl Genet 97 : 264–272
Sneath PHA, Sokal RR (1973) Numerical taxonomy. W.H. Freeman and Co., San Francisco
Szewc-McFadden AK, Bleik S, Alpha CG, Lamboy WF, McFerson JR (1995) Identification of simple sequence repeats in *Malus* (apple). HortScience 30 : 855
Szewc-McFadden AK, Lamboy WF, Hokanson SC, McFerson JR (1996) Utilization of identified simple sequence repeats (SSRs) in *Malus* × *domestica* (apple) for germplasm characterization. HortScience 31 : 619
Tanksley SD, McCouch SR (1997) Seed banks and molecular maps: unlocking genetic potential from the wild. Science 277 : 1063–1066
Tanksley SD, Nelson JC (1996) Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor Appl Genet 92 : 191–203
Tanksley SD, Grandillo S, Fulton TM, Zamir D, Eshed Y, Petiard V, Lopez J, Beck-Bunn T (1996) Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative *L. pimpinellifolium*. Theor Appl Genet 92 : 213–224
Thomas MR, Scott NS (1993) Microsatellite repeats in grapevine reveal DNA polymorphisms when analysed as sequence-tagged sites (STSs). Theor Appl Genet 86 : 985–990
Ven WTG van de, McNicol RJ (1996) Microsatellites as DNA markers in Sitka spruce. Theor Appl Genet 93 : 613–617
Weir BS (1989) Sampling properties of gene diversity. In: Brown AHD, Clegg MT, Kahler HL, Weir BS (eds) Plant population genetics, breeding and genetic resources. Sinauer Associates, Sunderland, Massachusetts, pp 23–42
Wu KS, Tanksley SD (1993) Abundance, polymorphism and genetic mapping of microsatellites in rice. Mol Gen Genet 241 : 225–235
Xiao J, Grandillo S, Ahn SN, McCouch SR, Tanksley SD (1996) Genes from wild rice improve yield. Nature 384 : 223–224
Zimmerer KS, Douches DS (1991) Geographical approaches to crop conservation: the partitioning of genetic diversity in Andean potatoes. Econ Bot 45 : 176–189