Identification of “Duplicate” Accessions within the USDA-ARS National Plant Germplasm System *Malus* Collection

Briana L. Gross, Gayle M. Volk², and Christopher M. Richards

USDA-ARS National Center for Genetic Resources Preservation, 1111 S. Mason Street, Fort Collins, CO 80521

Philip L. Forsline¹, Gennaro Fazio, and C. Thomas Chao

USDA-ARS Plant Genetic Resources Unit, Geneva, NY 14456-0462

ADDITIONAL INDEX WORDS. apple, *Malus ×domestica*, simple sequence repeat, probability of identity, clonal crop, perennial crop

ABSTRACT. The U.S. Department of Agriculture, Agricultural Research Service, National Plant Germplasm System (NPGS), Plant Genetic Resources Unit apple (*Malus*) collection in Geneva, NY, conserves over 2500 trees as grafted clones. We have compared the genotypes of 1131 diploid *Malus ×domestica* cultivars with a total of 1910 wild and domesticated samples representing 41 taxonomic designations in the NPGS collection to identify those that are genetically identical based on nine simple sequence repeat (SSR) loci. We calculated the probability of identity for samples in the data set based on allelic diversity and, where possible, use fruit images to qualitatively confirm similarities. A total of 237 alleles were amplified and the nine SSRs were deemed adequate to assess duplication within the collection with the caveat that “sport families” likely would not be differentiated. A total of 238 *M. ×domestica* and 10 samples of other taxonomic groups shared a genotype with at least one other *M. ×domestica* individual. In several cases, genotypes for cultivars matched genotypes of known rootstocks and indicated that these accessions may not accurately represent the indicated named clones. Sets of individuals with identical genotypes and similar cultivar names were assigned to sport families. These 23 sport families, comprised of 104 individuals, may have mutational differences that were not identified using the nine SSR loci. Five of the selected markers (CH01h01, CH02d08, CH01f02, G12, GD147) overlap with sets of markers that have been used to fingerprint European apple collections, thus making it possible to compare and coordinate collection inventories on a worldwide scale.

The global production of apples is threatened by disease, pest susceptibility, suboptimal cold-hardiness and heat tolerance, minimal resistance to drought and wet soils, undesirable storage and transport characteristics, and expensive production methods. The U.S. Department of Agriculture, Agricultural Research Service (USDA-ARS) National Plant Germplasm System, Plant Genetic Resources Unit (PGRU) apple collection in Geneva, NY, conserves key genetic resources useful for breeding and research programs that address threats to apple crop production. The *Malus* collection is maintained as grafted, clonally propagated trees and own-rooted seedling trees. The grafted orchards include 33 species of *Malus*, of which only one species, *Malus ×domestica*, is cultivated for fruit production, as well as some hybrid materials. These species are represented by between one and 1372 unique accessions each. The taxonomic and ecological breadth of this collection makes it a vital genetic resource for both the United States and the world.

The clonally propagated apple field collection is grafted on ‘East Malling 7’ (EMLA 7) rootstock (Wertheim, 1998), which was selected for its dwarfing and increased resistance to

fireblight infection (Aldwinckle et al., 2004; Forsline et al., 2010). Currently, the grafted collection has ≈2500 clones acquired primarily from gene banks, breeders, and wild collection trips (Table 1). The field collection also includes several thousand *Malus* species seedling (non-grafted) trees, derived primarily from seeds collected during plant exploration trips. Field trees are expensive to maintain; the most recent estimated cost was more than \$50 per tree each year (Hokanson et al., 1998), and this figure has almost certainly been exceeded in recent years. Thus, it is critical to understand the genetic constitution of the collection materials to ensure continued maintenance of high-priority genetic resources using limited financial resources.

Microsatellite markers have been used successfully to assess the diversity of wild and clonally propagated fruit collections (Gökirmak et al., 2009; Koehmstedt et al., 2011; Laucou et al., 2011; Njuguna et al., 2011; Urrestarazu et al., 2012; van Treuren et al., 2010). As a result of the high levels of allelic diversity within most *Malus* species, relatively few markers are required to differentiate among unique *M. ×domestica* cultivars (Hokanson et al., 1998) or among half-sib individuals from wild populations of *M. sieversii* and *M. orientalis* (Richards et al., 2009a, 2009b; Volk et al., 2005, 2008). Conversely, the very low genomic coverage of microsatellite data sets limits our ability to differentiate among individuals in “sport” families, because these individuals may exhibit distinct phenotypic traits that are based on one or a few mutations (Venturi et al., 2006; Wünsch and Hormaza, 2002).

In this work, we determined whether a set of nine SSR loci is sufficiently variable to differentiate among siblings and identify

Received for publication 16 May 2012. Accepted for publication 17 July 2012. We appreciate the technical support provided for this project by Adam D. Henk, Sarah Bauer, Amy K. Szevc-McFadden, and Celeste Falcon. Lee Panella provided a critical manuscript review before publication.

Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Dept. of Agriculture.

¹Retired.

²Corresponding author. E-mail: Gayle.Volk@ars.usda.gov.

Table 1. Taxonomic designation and number of individual inventories of *Malus* trees in the Geneva, NY, Plant Genetic Resources Unit (PGRU) grafted orchard, the number of those grafted trees that are known to be diploid, the number of individuals included in the nine simple sequence repeat (SSR) data set, and the number of seedling trees in the nine SSR data set.

Species	PGRU orchard		9 SSR data set	
	Grafted clones (no.)	Diploid grafted clones (no.)	Grafted clones (no.)	Seedlings (no.)
<i>Malus angustifolia</i>	17	7	0	
<i>Malus asiatica</i>	16	10	7	
<i>Malus baccata</i>	51	47	42	
<i>Malus bhutanica</i>	7	1	1	
<i>Malus brevipes</i>	2	2	2	
<i>Malus coronaria</i>	40	4	3	
<i>Malus ×domestica</i>	1372	1227	1131	
<i>Malus florentina</i>	3	3	0	
<i>Malus floribunda</i>	10	10	8	
<i>Malus fusca</i>	40	40	8	
<i>Malus halliana</i>	14	8	8	
<i>Malus hupehensis</i>	16	3	2	
<i>Malus hybrid</i> ^z	328	317	235	
<i>Malus ioensis</i>	36	17	0	
<i>Malus kansuensis</i>	6	6	3	26
<i>Malus kirghisorum</i>	15	15	7	
<i>Malus mandshurica</i>	3	3	3	
<i>Malus micromalus</i>	15	13	9	
<i>Malus ombrophila</i>	3	3	0	
<i>Malus orientalis</i>	16	16	4	45
<i>Malus orthocarpa</i>	1	1	1	
<i>Malus prattii</i>	4	4	0	4
<i>Malus prunifolia</i>	36	35	29	
<i>Malus pumila</i>	12	11	8	
<i>Malus sargentii</i>	18	1	1	
<i>Malus sieversii</i>	283	283	151	1
<i>Malus sikkimensis</i>	13	2	2	
<i>Malus sp.</i> ^y	29	23	17	
<i>Malus spectabilis</i>	7	5	2	
<i>Malus sylvestris</i>	22	21	17	43
<i>Malus toringo</i>	22	12	9	
<i>Malus transitoria</i>	4	2	2	17
<i>Malus tschonoskii</i>	3	3	0	
<i>Malus x</i> ^x	51	44	36	
<i>Malus yunnanensis</i>	15	13	5	
<i>Malus zhaojiaoensis</i>	2	2	2	19
Total	2532	2214	1755	155

^z*Malus* hybrid indicates that the accession appears to be a hybrid between two species, but its parentage is not known.

^y*Malus* sp. indicates that the accessions are likely not *Malus ×domestica* (based on phenotypic differences), but they have not been positively identified at the species level.

^x*Malus* × includes *M. ×adstringens*, *M. ×arnoldiana*, *M. ×atrosanguinea*, *M. ×dawsoniana*, *M. ×hartwigii*, *M. ×magdeburgensis*, *M. ×moerlandsii*, *M. ×platycarpa*, *M. ×purpurea*, *M. ×robusta*, *M. ×scheideckeri*, *M. ×soulardii*, *M. ×sublobata*, and *M. ×zumi*.

identical accessions in the USDA-ARS *Malus* field collection based on the probability of identity (ProbI) for the data set. We identified genetically identical *M. ×domestica* cultivars maintained as clones in the field collection and, where possible, used fruit images to qualitatively confirm similarities.

Materials and Methods

PLANT MATERIAL. Most of the trees in the grafted PGRU orchards were sampled for potential inclusion in the genetic analyses, and, in some cases, seedling trees were included as additional representatives of wild species. For both grafted and seedling trees, only diploid trees had microsatellite signatures

that could be scored in a comparative manner, and only these were included in the final data set. The final data set included 1910 individuals, or accessions, each identified with a unique Plant Introduction number [PI# (for grafted trees)] or Geneva *Malus* number (for seedlings in most cases) identification numbers (Table 1). For the present analyses, we focus on 1131 diploid *M. ×domestica* cultivars in the PGRU orchards. To identify duplicates, each *M. ×domestica* genotype has been compared with all the other domesticated genotypes as well as 779 wild genotypes in the 1910 individual data set.

MICROSATELLITE MARKERS. Genomic DNA was extracted from leaf tissue using DNeasy 96 plant kits (Qiagen, Valencia, CA). Nine previously published SSRs [GD12, GD15, GD96,

GD142, GD147, GD162, CH01h01, CH01f02, and CH02d08 (Hokanson et al., 1998; Liebhard et al., 2002)] chosen based on their ease of scoring, variation, and overlap with SSRs used to evaluate *Malus* collections in other countries were amplified in all samples. Primer sequences, amplicon size ranges, and annealing temperatures are listed in Gross et al. (2012). Forward primers, labeled with either IRD 700 or IRD 800, were obtained from MWG-Biotech (High Point, NC). Unlabeled reverse primers were purchased from IDT (Coralville, IA). All polymerase chain reactions (PCRs) were carried out in 15 μ L total volume using previously published methods (Volk et al., 2005). PCR products were visualized and scored in one of two ways. Some were visualized on a slab sequencer (LI-COR 4200; LI-COR, Lincoln, NE); digital images were collected from the sequencer using LI-COR Saga Generation 2 software and were manually interpreted and scored using the Saga software. Others were visualized on a capillary sequencer (ABI 3730; Applied Biosystems, Foster City, CA); chromatograms were scored using the GeneMarker software (SoftGenetics, State College, PA). Peaks were scored automatically based on allelic bins created for each SSR locus and then corrected manually for cases of errors or nonsensical automatic scoring. When a single SSR was scored using both the LI-COR and ABI systems, a minimum of 10 individuals were run on both systems to control for allele length differences resulting from instrumentation.

PROBABILITY MEASUREMENTS AND IDENTIFICATION OF "DUPLICATES." Basic diversity statistics, along with polymorphism information content and discriminatory power for each locus in the nine SSR data set, were calculated using GenoDive and CERVUS (Botstein et al., 1980; Kalinowski et al., 2007; Meirmans and Van Tienderen, 2004; Tessier et al., 1999). Note that expected heterozygosity (H_e) and observed heterozygosity (H_o) were calculated for reference but are not comparable to those calculated for a randomly mating population. To evaluate our ability to correctly identify duplicated genotypes using the nine SSR data set, we calculated ProBI for each locus and across multilocus genotypes. ProBI is an estimate of the probability that two individuals will have the same multilocus genotype by chance alone (Peakall and Smouse, 2006; Taberlet and Luikart, 1999). To account for relatedness among samples, we also calculated the more conservative ProBI_{sibs}, which allows for the possibility of identity resulting from sibling relationships. Both ProBI and ProBI_{sibs} were calculated using GenAlEx (Peakall and Smouse, 2006). The nine SSR data set, with data missing at a maximum of one locus for any individual, was checked for duplicate genotypes using GenoDive (Meirmans and Van Tienderen, 2004). Genotypes were considered to be duplicated if they matched at all alleles across all nine SSRs with the exception of missing data, which were ignored (i.e., counted neither for nor against identity).

PHENOTYPES AND FRUIT IMAGES. Fruit images for each accession identified as a duplicate were downloaded from the Germplasm Resources Information Network database (USDA, 2012). Images were qualitatively compared for fruit size, fruit ground color, fruit over color, shape, and russet appearance.

Results

This research focuses on the *M. $\times$ domestica* cultivars in the grafted collection but does include some seedling materials for comparison purposes. The *M. $\times$ domestica* collection has 1372 individuals, of which 1240 (90%) are diploid, and thus produced amplification products that could be accurately scored. Individuals with missing data at more than one locus were not included in the data set, leaving 1131 *M. $\times$ domestica* accessions that were compared with each other as well as the additional 779 diploid individuals from other species in the data set.

Nine SSRs were amplified in the 1910 individuals, and a total of 237 alleles (average of 26.3 alleles per locus) were

Table 2. Genetic parameters for each of the nine simple sequence repeat (SSR) loci used to determine duplicate genotypes in the *Malus* collection, including number of alleles (A), observed heterozygosity (H_o), expected heterozygosity (H_e), the probability of identity (ProBI), the discriminating power (D), and the polymorphic information content (PIC).

SSR locus	A	H_o	H_e	ProBI	D	PIC
GD12	31	0.66	0.89	0.054	0.946	0.788
GD15	5	0.37	0.50	0.601	0.399	0.221
GD96	31	0.79	0.94	0.016	0.984	0.898
GD142	26	0.81	0.91	0.018	0.982	0.894
GD147	26	0.55	0.93	0.027	0.973	0.864
GD162	30	0.66	0.94	0.026	0.974	0.866
CH01h01	28	0.74	0.94	0.020	0.980	0.887
CH01f02	34	0.86	0.93	0.013	0.987	0.910
CH02d08	26	0.67	0.90	0.030	0.970	0.854

Fig. 1. Probabilities of identity (ProBI) for increasing combination of the nine simple sequence repeat (SSR) loci arranged in decreasing order of ProBI for each locus. Dashed lines show the standard ProBI, and solid lines show the more conservative ProBI_{sibs}.

present across the nine SSRs. Levels of heterozygosity were high; H_o was greater than 0.50 and H_e was greater than 0.85 across all species for all loci except for GD15 (Table 2). In *M. ×domestica*, H_o was greater than 0.70 for all loci except for GD15 (data not shown). $ProbI$ and $ProbI_{sibs}$ were calculated for the data set to determine whether any empirically determined duplicated genotypes were likely the result of chance alone or whether they were the result of true identity at the genetic level. Although there is not a firm cutoff for when to interpret a level of $ProbI$ as sufficient, one suggested standard (Peakall et al., 2006) is the reciprocal of the sample size ($1/1910 = 0.00052$). Thus, a value for $ProbI$ or $ProbI_{sibs} < 0.00052$ would indicate that when samples do share identical genotypes, it is the result of true identity between the individuals. The values of $ProbI$ and $ProbI_{sibs}$ are plotted in Figure 1, and the required level of $ProbI$ is achieved with two SSRs for $ProbI$ and with seven SSRs for the more conservative $ProbI_{sibs}$. $ProbI$ values for each SSR

locus, representing a measure of the discriminatory power of a given marker, are given in Table 2.

Given that the number and variability at the loci used in this study are likely sufficient to discriminate between individuals and matches that represent true identity, another way to interpret the data are to consider the probability (however low) that two individuals would match by chance alone. For the standard value of $ProbI$, assuming no relatedness, the probability is 4.8×10^{-15} , or one in 21 trillion. For the more conservative value of $ProbI_{sibs}$, it is 7.7×10^{-5} , or one in 13,000. The real value is likely somewhere in between these two extremes, because some of the individuals are undoubtedly related, but many are not, given the number of different species present in the data set.

The explicit possibility that full- or half-sibs might have the same genotype based on their close relatedness also must be considered, especially in the context of the sport families. The probability that two full-sibs will have the same genotype is

Fig. 2. Chromatograms of locus CH02d08 for two sets of duplicated *Malus × domestica* genotypes. The first two are PI #437036 and PI #613830 (cultivars Aromat de Vara and Inducoa No. 11) with alleles of length 212 and 230 bps. The second two are PI #129820 and PI #249925 (cultivars Cole and King Cole) with alleles of length 214 and 230 bps.

Table 3. Plant Introduction (PI) or Geneva *Malus* (GMAL) number, cultivar name (followed by species name if not *Malus ×domestica*), and country of origin for matching genotypes or sets of genotypes, arranged horizontally.^z

PI/GMAL no.	Cultivar name	Country of origin ^z	PI/GMAL no.	Cultivar name	Country of origin
Same genotype, similar images					
PI 344547	Alkmene		PI 105498	Bella de Jardins	France
PI 187296	Bouteille	Germany	PI 158731	Bramtot	United Kingdom
PI 162549	Boche	France	PI 188524	Storappel	The Netherlands
PI 162545	Grosse Mouche	France	PI 589288	Not named (hybrid)	
PI 589593	Blenheim Orange	United Kingdom	PI 175551	Twistbody Jersey	United Kingdom
PI 347268	Kokko strain 2	Japan	PI 437047	Red Ralls	Poland
PI 589690	Le Bret	United Kingdom	PI 589081	Sweet Alford	United Kingdom
PI 589196	Crow Egg	United States	PI 589123	Geeveston Fanny	Australia
PI 589555	Joys	Former Soviet Union	PI 589444	Reinette Clochard	France
PI 589354	Cheal's Weeping		PI 78170	Oekonomierat Echter-meyer	Germany
PI 131215	Nanny	United Kingdom	PI 137056	Surprise Reinette	Switzerland
PI 161840	Lorna Doone	United Kingdom	PI 205460	Merton Beauty	United Kingdom
PI 437036	Aromat de Vara	Romania	PI 175010	King Harry	United Kingdom
PI 613830	Inducoa No. II	Spain	PI 589153	Magnolia Gold	United States
PI 589135	Golden Noble	United Kingdom	PI 188525	Vogelcalville	The Netherlands
PI 349324	Coast Apple	South Africa	PI 209939	Subtropical Apple	South Africa
PI 276567	Orleans	United States	PI 589303	Quindell	United States
PI 279643	Ivette	The Netherlands	PI 589502	Z 71	The Netherlands
PI 589254	Doud Golden Delicious 2-4-4-4	United States	PI 283698	Imperial Stayman	United States
PI 383505	Ermak	former Soviet Union	PI 589908	Whitney Crab (hybrid)	United States
PI 589215	Crimson Gold (pumila)	United States	PI 437064	Yar Mohammadi	Pakistan
PI 537000	Drakenstein	South Africa	PI 589342	E 12-14	South Africa
PI 589446	Idajon	United States	PI 588793	Snow	Canada
PI 589574	Borowitsky	Russia	PI 588801	Duchess	United Kingdom
PI 588841	Idared	United States	PI 589553	Mantuanskoye	Italy
PI 589150	Lowry	United States	PI 589001	Winter Sweet Paradise	
Same genotype, image, and name					
PI 588900	Jubilee (hybrid)	Canada	PI 589725	Jubilee	Canada
PI 279645	Odin	The Netherlands	PI 589921	Odin ^y	The Netherlands
PI 280029	Pitmaston Pine Apple (hybrid)	United Kingdom	PI 279323	Pitmaston Pineapple × 692	France
PI 589457	Rosu de Cluj	Romania	PI 613831	Rosu de Cluj	Romania
PI 588849	Russian	Europe	PI 589312	Russian sdlg.	
PI 589025	Splendor (Stark)	New Zealand	PI 437054	Splendour (hybrid)	New Zealand
PI 483256	YP	Finland	PI 588827	YP (MB4) (pumila)	Finland
Same genotype, paired images not available					
PI 589540	Alpinist	Former Soviet Union	PI 589560	Livadiyskoye	Ukraine
PI 267939	Ananas Berzenicki	Poland	PI 262962	Prof. Grebnicka Renete	Latvia
PI 589885	Beacon	United States	PI 590171	Northern Lights	
PI 589664	Beauty of Bath	United Kingdom	PI 102140	Gladstone	United Kingdom
PI 126495	Belle Imperiale	United Kingdom			
GMAL 4662	Bisquet	France	PI 589670	Michelin ^x	France
PI 122616	Doux-AMR				
PI 589290	C-13-30-88 (hybrid)		PI 589504	Kestrel	Canada
PI 383503	Chulanka	former Soviet Union	PI 383504	Creamy Kitaika	Former Soviet Union
GMAL 4660	Clos Renaux	France	PI 589689	Frequin Tardive de la Sarthe	United Kingdom
PI 589191	Collamer Twenty Ounce	United States	PI 589007	Twenty Ounce ^y	United States
PI 589149	Collins June	United States	PI 589015	Early Harvest	United States
PI 589293	Early Banta	United States	PI 589847	Holland	United States
PI 392312	Edward VII	United Kingdom	PI 589678	King Edward VII	United Kingdom
PI 157207	Gold Canel	Finland	PI 157211	Red Canel	Finland
PI 589491	Korichnoe Polosatoje	former Soviet Union	PI 383520	Striped Tawny	Former Soviet Union
PI 590220	Hardy Cumberland	United States	PI 588773	Yellow Newtown	United States
PI 136001	Herring's Pippin	United Kingdom	PI 260002	San Juan	Spain
PI 589054	Red Astrachan	Russia			
PI 589598	La Paix	Belgium	PI 589077	Mother	United States

continued next page

Table 3. Continued.

PI/GMAL no.	Cultivar name	Country of origin ^z	PI/GMAL no.	Cultivar name	Country of origin
PI 589646	Laxton's Fortune	United Kingdom	PI 127009	Laxton's Fortune	United Kingdom
PI 589367	Mantet	Canada	PI 589466	Roda Mantet	The Netherlands
PI 264558	Margil	United Kingdom	PI 279325	Reine des Reinettes × 82	France
PI 173983	Metais	France	PI 162735	Mottais	France
PI 259451	Michurin's Bessiemianka	Former Soviet Union	PI 588964	Michurin's Seedless	
PI 589719	Ottawa 274	Canada	PI 589340	Ottawa 292	Canada
PI 589323	Paradiso		PI 589021	Young America	United States
PI 588821	Patul	Romania	PI 175552	Vagnon Ascher	United Kingdom
PI 589787	PRI 1033-5 'Viking' (hybrid)	United States	PI 589434	Viking	United States
PI 105524	Reineta do Caravia	Spain	PI 245145	Toreno	Spain
PI 589561	Shaphran Letnij	Former Soviet Union	PI 589550	Slava Pobeditelyam	Ukraine
PI 104815	Suislepper	Sweden	PI 292931	Susvorenskoye	Former Soviet Union
PI 589433	Trebu		PI 262966	Trebu Seklandzis	Latvia
PI 588855	Tydemman Red (Cooper)		PI 206035	Worcester Pearmain	United Kingdom
PI 241965	Tydemans Early Worcester	United Kingdom			
PI 589558	Tytovka	Former Soviet Union	PI 126998	Zigeunerinapfel	The Netherlands
PI 589205	Wrixparent 2-4-4-4	United States	PI 588859	Yellow Transparent	Russia
Rootstock					
PI 263633	Malus sp. 'Rootstock'	United Kingdom	PI 589710	Merton 793	
PI 437058	Malling III (M.13)	United Kingdom	PI 589928	M.13	The Netherlands
PI 588812	Malling Merton 111 (MM.111)	United Kingdom	PI 613944	Coop 44 (sp.)	United States
PI 589923	Red Dijmanszoet	The Netherlands			

^zEach unshaded row represents a pair of genotypes that match. Sets of grayed rows represent a group of three or more matched genotypes.

^yCountry of origin is based on information in the Germplasm Resources Information Network (GRIN) database (U.S. Department of Agriculture, 2012) and sometimes refers to the country from which the accession was received.

^xUnavailable in 2012.

mainly dependent on the level of homozygosity of the parents and ranges from 0.25 if the parents are heterozygous with four different alleles across a single locus to 1.0 if the parents are homozygous. There are a variety of ways that these probabilities can be combined across nine SSRs. However, even if one or the other parent is homozygous at each of the nine SSRs (although the other is heterozygous), the probability that the siblings would have identical genotypes would be 0.00195, or less than one in 500. Given the high levels of observed heterozygosity for most loci in the data set (H_o greater than 0.70 for eight of the nine loci in *M. ×domestica*), it is unlikely that two parents would have this many homozygous loci. Thus, any matching genotypes are much more likely to represent duplicates than full-sibs.

Overall, based on the allelic diversity, levels of heterozygosity, and ProBI and $\text{ProBI}_{\text{sibs}}$ values for the data set, the nine SSRs were deemed adequate to assess duplication within the collection with the caveat that "sport families" likely would not be differentiated. A total of 238 *M. ×domestica* and 10 samples of other taxonomic groups (eight *Malus* hybrid, one *M. pumila*, one *Malus* species) shared a genotype with at least one other *M. ×domestica* individual (Fig. 2). When available, fruit image data generally supported the genetically similar assessments based on the nine SSR loci (Table 3; Fig. 3). Fifty-two cultivars with unique names shared 24 genotypes and their similarities could be confirmed visually with image data for those individuals. Seven pairs of cultivars with very similar names also shared image and genotypic identities. Seventy individuals shared 32 genotypes, but image data were not available (Table 3).

The data set included two rootstock accessions named 'Malling XIII' (M.13) with identical genotypes. The genotype of PI #589710 'Merton 793' matched that of PI #263633 *Malus* sp. 'Rootstock', and the genotype of PI #589923 'Red Dijmanszoet'

and PI #613944 'Coop 44' (sp.) matched PI #588812 'Malling Merton 111' (MM.111), another known rootstock. Based on the identical genetic signatures between these "cultivars" and rootstocks, it appears as if these specific accessions may not represent accurately the indicated named clones (Table 3). Instead, these individuals may have been received as materials believed to be clones, although they were actually rootstocks at the original source planting. When sets of individuals were identified as being genetically identical, had related cultivar names, and mostly similar fruit appearances, they were assigned to "sport" families. There were 23 sport families containing between two and 15 individuals each for a total of 105 individuals in all (Table 4). These sport families may have mutational differences that could not be identified using the nine SSR loci.

Discussion

The USDA seeks to maintain a diverse collection of *Malus* genetic resources. Given the maintenance cost per tree and the limited space available, it is critical that new materials strategically expand the collection by capturing novel genetic diversity that has potential value to the user community. Effectively maintaining the genetic diversity of clonal crop collections such as the NPGS *Malus* orchard is particularly important in the context of modern production, in which the occasional infusions of genetic diversity through cross-pollination with wild relatives, once a major source of new variation, are rarely incorporated into crop germplasm (McKey et al., 2010). Indeed, the U.S. apple industry is based primarily on 11 apple cultivars (Dennis, 2008), and these modern commercial cultivars are derived from progeny of only four seedling parents: 'Cox's Orange Pippin', 'Golden Delicious', 'Jonathan', and 'McIntosh'

Fig. 3. Representative sets of *Malus x domestica* fruit identified sharing identical genotypes based on nine simple sequence repeat (SSR) loci. PI number and cultivar names are provided on the figure for pairs of accessions with matching genotypes and different names (A–D), cultivars matching rootstock accessions (E–F), and a portion of the ‘Golden Delicious’ sport family (G). Fruit are arranged on 1-cm² grids.

(Brown, 2012; Luby, 2003; Noiton and Alspach, 1996). Diverse, well-characterized apple collections of both *M. x domestica* as well as wild *Malus* species provide breeding programs with novel alleles that have the potential to increase yield, value, and quality as well as decrease susceptibility to biotic and abiotic stresses. Essentially, these collections serve as a resource to maintain a high overall genetic diversity for a crop despite very low diversity in the field.

Genetic assessments of the USDA-ARS-NCGRP apple collection allow for a more in-depth understanding of diversity

represented by the greater than 1000 diploid *M. x domestica* trees currently maintained as part of the orchard. This research focuses on identifying and classifying the “duplicate” materials within the collection. Accurately identifying genetically identical samples in a data set of any sort (crop collections, mark-recapture studies, forensics) requires that the selected loci be sufficiently variable and that the calculated probability of two genotypes matching take into account the reality of population substructure and relatedness among individuals to avoid matches that are the result of relatedness alone (Peakall et al., 2006; Waits et al., 2001). We have dealt with these issues by calculating both ProbI and $\text{ProbI}_{\text{sibs}}$. The latter measure serves as a “conservative upper bound for the probability of observing identical multilocus genotypes between two individuals sampled from a population” (Waits et al., 2001). The large number of alleles amplified at the SSR loci used in this study (with the exception of GD15), combined with high levels of heterozygosity in wild and domesticated apple, yielded levels of ProbI and $\text{ProbI}_{\text{sibs}}$ that were low enough to be considered sufficient for both the size of the data set (0.00052) and the recommended levels of ProbI [0.01–0.0001 (Waits et al., 2001)]. Thus, the “duplicates” identified by this genetic analysis (and confirmed, in some cases, by fruit images) are likely to be very closely related, if not actually identical.

With “duplicates” identified in studies like this one, curators can assess if likely misidentified materials (such as the cultivars that match rootstocks) should be maintained or removed to accommodate new samples. In addition, materials identified as identical based on the nine SSRs can be characterized

further at the phenotypic level to determine if the cultivars possess unique traits or if they might be similar enough to be considered synonyms of genotypes in the collection. This approach can also be taken with the sets of sport families, which can be characterized for phenotypic or physiological traits of interest. Sport family members may have unique, desirable traits that are based on simple mutations (potentially as a result of retrotransposon activity) compared with other family members (Kobayashi et al., 2004; Sun et al., 2010; Venturi et al., 2006). The cultivar Wijcik McIntosh is an

Table 4. Sport family name, PI number, cultivar name, and country of origin for matching genotypes or sets of genotypes, arranged horizontally.

Sport family	PI no.	Cultivar name	Country of origin ^z
Antonovka	PI 588784	Antonovka	Russia
Antonovka	PI 589837	Antonovka Monasir	
Antonovka	PI 589329	Antonovka Polutorafuntovaya	Former Soviet Union
Antonovka	PI 322715	Tayeshnoe	Former Soviet Union
Belflower	PI 589577	Brabant Bellefleur	The Netherlands
Belflower	PI 264692	Red Belflower	Serbia
Ben Davis	PI 590025	Anisim	Former Soviet Union
Ben Davis	PI 588953	Ben Davis	United States
Ben Davis	PI 589178	Black Ben Davis	United States
Cheddar	PI 199100	Cheddar	United Kingdom
Cheddar	PI 589656	Cheddar Cross	United Kingdom
Cole	PI 129820	Cole	United Kingdom
Cole	PI 249925	King Cole	Australia
Court Pendu	PI 589601	Court Pendu	Belgium
Court Pendu	PI 589602	Court Pendu Gris ^y	France
Cox	PI 247022	Cox's Orange Cherry	Denmark
Cox	PI 247023	Cox's Orange Otago	
Cox	PI 588853	Cox's Orange Pippin	United Kingdom
Cox	PI 589127	Potter Cox	United Kingdom
Golden Delicious	PI 589139	Badami Golden Delicious	United States
Golden Delicious	PI 589224	Clear Gold	United States
Golden Delicious	PI 589316	Empress Spur Golden Delicious	United States
Golden Delicious	PI 590184	Golden Delicious	United States
Golden Delicious	PI 589041	Golden Delicious (Smoothee)	United States
Golden Delicious	PI 589535	Golden Delicious SE-69	Czech Republic
Golden Delicious	PI 589448	Golden Precoce	France
Golden Delicious	PI 644190	Goldspur Golden Delicious	United States
Golden Delicious	PI 589966	Puregold	United States
Golden Delicious	PI 589136	Razor Golden Delicious	United States
Golden Delicious	PI 589125	Sergeant Russet Golden Delicious	United States
Golden Delicious	PI 589904	Smoothgold	
Golden Delicious	PI 589210	Starkspur Golden Delicious	United States
Golden Delicious	PI 589528	Woodward	United States
Golden Delicious	PI 589856	Yellow Delicious	United States
Jersey	PI 161845	Red Jersey	United Kingdom
Jersey	PI 175545	Royal Jersey	United Kingdom
Jonathan	PI 194199	Bowden Seedling	United Kingdom
Jonathan	PI 589160	Conkle Jonathan 2-4-4-4	United States
Jonathan	PI 588940	Jonared	United States
Jonathan	PI 590185	Jonathan	United States
Jonathan	PI 590150	Kingjon	United States
Jonathan	PI 589097	Welday Jonathan 2-2-4-4	United States
Jonathan	PI 589144	Welday Jonathan 4X	United States
Lamb Abbey Pearmain	PI 199418	Lamb Abbey Pearmain-A	United Kingdom
Lamb Abbey Pearmain	PI 199419	Lamb Abbey Pearmain-B	United Kingdom
McIntosh	PI 588961	Boller McIntosh	United States
McIntosh	PI 589557	Calvil Crymskiy	Former Soviet Union
McIntosh	PI 589183	Cornwall McIntosh 2-2-2-4	United States
McIntosh	PI 589069	Green Peak McIntosh 2-2-4-4	United States
McIntosh	PI 589507	Hamilton (hybrid)	Canada
McIntosh	PI 589206	Johnson McIntosh 2-4-4-4	United States
McIntosh	PI 589122	Kimball McIntosh 2-4-4-4	United States
McIntosh	PI 588999	MacSpur	Canada
McIntosh	PI 588998	Marshall McIntosh	United States
McIntosh	PI 588817	McIntosh Summerland Red	Canada
McIntosh	PI 588968	Rogers McIntosh	United States
McIntosh	PI 590186	Wijcik McIntosh	Canada
Merton	PI 589849	Merton Ace	United Kingdom

continued next page

Table 4. Continued.

Sport family	PI no.	Cultivar name	Country of origin ^z
Merton	PI 589429	Merton Knave	United Kingdom
Muscadet	PI 162544	Belle de Crollon	France
Muscadet	PI 200780	Muscadet Bernay	France
Muscadet	PI 173985	Muscadet de Lense	France
Orange	PI 589575	Ellison's Orange	United Kingdom
Orange	PI 392307	Finson's Orange	Australia
Paragon	PI 589193	Dermen Paragon 3-6-6	United States
Paragon	PI 589148	Dermen Paragon 6-3-3	United States
Paragon	PI 589039	Paragon	United States
Red Delicious	PI 589841	Delicious	United States
Red Delicious	PI 589110	Dermen Delicious 4X	United States
Red Delicious	PI 589068	Earlibrite Delicious	United States
Red Delicious	PI 589064	Fruitland Delicious	United States
Red Delicious	PI 590026	Hawkeye	United States
Red Delicious	PI 589037	Idaho Spur Delicious	United States
Red Delicious	PI 589255	Redspur Delicious	United States
Red Delicious	PI 589151	Richared Delicious	United States
Red Delicious	PI 589137	Rose Red Delicious	United States
Red Delicious	PI 589192	Starking	United States
Red Delicious	PI 589221	Yellow Red Delicious	United States
Rome	PI 589201	Barkley Rome	United States
Rome	PI 589034	Gallia Beauty	United States
Rome	PI 588986	Hotle Rome	United States
Rome	PI 589033	Loop Rome 2-2-2-4	United States
Rome	PI 588956	Milton	United States
Rome	PI 588850	Rome Beauty Law	United States
Spartan	PI 589231	Hunter Spartan 2-4-4	Canada
Spartan	PI 588871	Spartan	Canada
Spartan	PI 589204	Sweden Spartan	Sweden
Spitzenburg	PI 588785	Esopus Spitzenburg	United States
Spitzenburg	PI 589100	Red Spitzenburg	United States
Spy	PI 589047	Farmer Spy	United States
Spy	PI 589055	Field Spy	United States
Spy	PI 589132	Green Peak Spy #1	
Spy	PI 589203	Hunter Spy 2-4-4	Canada
Spy	PI 589207	Loop Giant Spy	United States
Spy	PI 589096	Loop Red Spy	United States
Spy	PI 589027	Loop Spy 2-2-4-4	
Spy	PI 588872	Northern Spy	United States
Spy	PI 589842	Red Spy	
Spy	PI 588969	Schoharie Spy	United States
Stayman	PI 589152	Dermen Black Stayman	United States
Stayman	PI 589236	Dermen Stayman 6-3-3	United States
Stayman	PI 588975	Stayman	United States
Wealthy	PI 589188	Coombs Wealthy	United States
Wealthy	PI 588788	Wealthy	United States
Wealthy	PI 588779	Wealthy Double Red PC-310	United States

^zCountry of origin is based on information in the Germplasm Resources Information Network (GRIN) database (U.S. Department of Agriculture, 2012) and sometimes refers to the country from which the accession was received.

^yUnavailable in 2012.

excellent example of such a sport; it exhibits a unique, heritable columnar growth habit. Sports like 'Wijcik McIntosh' with desirable traits could be selected for further genetic analyses to identify new loci controlling traits of interest or new alleles at known loci that could be useful for future breeding efforts.

This research presented data from nine SSR loci that are sufficient to determine potential duplicates among 1910 acces-

sions in the USDA *Malus* collection and likely from other large collections of *Malus* as well. These markers overlap those previously used to assess diversity within wild species (Richards et al., 2009a, 2009b; Volk et al., 2005, 2008, 2009) and among *M. ×domestica*, *M. orientalis*, *M. sieversii*, and *M. sylvestris* in the USDA-ARS-NPGS collection (Gross et al., 2012). Five of the markers (CH01h01, CH02d08, CH01f02,

G12, GD147) overlap with the markers that have been selected within the European community for collection comparison purposes (F. Laurens, personal communication; M. Ordidge, personal communication), so it will be possible to extend this type of analysis across collections in the future. Ultimately, studies of this type can be used to lay the groundwork for genomic analyses of apple diversity and functional loci.

Literature Cited

- Aldwinckle, H.S., N. LoGiudice, G. Fazio, J.L. Norelli, T.L. Robinson, H.T. Holleran, and W.C. Johnson. 2004. Resistance of apple rootstocks to fire blight infection caused by internal movement of *Erwinia amylovora* from scion infection. *Acta Hort.* 663:229–233.
- Botstein, D., R.L. White, M. Skolnick, and R.W. Davis. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Amer. J. Hum. Genet.* 32:314–331.
- Brown, S. 2012. Apple, p. 329–368. In: Badenes, M.L. and D.H. Byrne (eds.). *Fruit breeding*. Springer, New York, NY.
- Dennis, F.G., Jr. 2008. *Malus × domestica*: Apple, p. 661–674. In: Janick, J. and R.E. Paull (eds.). *The encyclopedia of fruits and nuts*. CABI Publishing, Cambridge, MA.
- Forsline, P.L., H.S. Aldwinckle, E.E. Dickson, J.J. Luby, and S.C. Hokanson. 2010. Collection, maintenance, characterization, and utilization of wild apples of central Asia. *Hort. Rev.* 29:1–61.
- Gökirmak, T., S. Mehlenbacher, and N. Bassil. 2009. Characterization of European hazelnut (*Corylus avellana*) cultivars using SSR markers. *Genet. Resources Crop Evol.* 56:147–172.
- Gross, B.L., A.D. Henk, P.L. Forsline, C.M. Richards, and G.M. Volk. 2012. Identification of interspecific hybrids among domesticated apple and its wild relatives. *Tree Genet. Genomes* (in press).
- Hokanson, S.C., A.K. Szewc-McFadden, W.F. Lamboy, and J.R. McFerson. 1998. Microsatellite (SSR) markers reveal genetic identities, genetic diversity and relationships in a *Malus × domestica* Borkh. core subset collection. *Theor. Appl. Genet.* 97:671–683.
- Kalinowski, S.T., M.L. Taper, and T.C. Marshall. 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol. Ecol.* 16:1099–1106.
- Kobayashi, S., N. Goto-Yamamoto, and H. Hirochika. 2004. Retrotransposon-induced mutations in grape skin color. *Science* 304:982.
- Koelmstedt, A., M. Aradhya, D. Soleri, J. Smith, and V. Polito. 2011. Molecular characterization of genetic diversity, structure, and differentiation in the olive (*Olea europaea* L.) germplasm collection of the United States Department of Agriculture. *Genet. Resources Crop Evol.* 58:519–531.
- Laucou, V., T. Lacombe, F. Dechesne, R. Siret, J.P. Bruno, M. Dessup, T. Dessup, P. Ortigosa, P. Parra, C. Roux, S. Santoni, D. Varès, J.P. Péros, J.M. Boursiquot, and P. This. 2011. High throughput analysis of grape genetic diversity as a tool for germplasm collection management. *Theor. Appl. Genet.* 122:1233–1245.
- Liebhart, R., L. Gianfranceschi, B. Koller, C.D. Ryder, R. Tarchini, E. Van De Weg, and C. Gessler. 2002. Development and characterisation of 140 new microsatellites in apple (*Malus × domestica* Borkh.). *Mol. Breed.* 10:217–241.
- Luby, J.J. 2003. Taxonomic classification and brief history, p. 1–14. In: Ferree, D.C. and I.J. Warrington (eds.). *Apples: Botany, production, and uses*. CABI Publishing, Cambridge, MA.
- McKey, D., M. Elias, B. Pujol, and A. Duputié. 2010. The evolutionary ecology of clonally propagated domesticated plants. *New Phytol.* 186:318–332.
- Meirmans, P.G. and P.H. Van Tienderen. 2004. GENOTYPE and GENODIVE: Two programs for the analysis of genetic diversity of asexual organisms. *Mol. Ecol. Notes* 4:792–794.
- Njuguna, W., K. Hummer, C. Richards, T. Davis, and N. Bassil. 2011. Genetic diversity of diploid Japanese strawberry species based on microsatellite markers. *Genet. Resources Crop Evol.* 58:1187–1198.
- Noiton, D.A.M. and P.A. Alspach. 1996. Founding clones, inbreeding, coancestry, and status number of modern apple cultivars. *J. Amer. Soc. Hort. Sci.* 121:773–782.
- Peakall, R., D. Ebert, R. Cunningham, and D. Lindenmayer. 2006. Mark–recapture by genetic tagging reveals restricted movements by bush rats (*Rattus fuscipes*) in a fragmented landscape. *J. Zool.* 268:207–216.
- Peakall, R. and P.E. Smouse. 2006. GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* 6:288–295.
- Richards, C.M., G.M. Volk, P.A. Reeves, A.A. Reilley, A.D. Henk, P.L. Forsline, and H.S. Aldwinckle. 2009a. Selection of stratified core sets representing wild apple (*Malus sieversii*). *J. Amer. Soc. Hort. Sci.* 134:228–235.
- Richards, C.M., G.M. Volk, A.A. Reilley, A.D. Henk, D. Lockwood, P.A. Reeves, and P.L. Forsline. 2009b. Genetic diversity and population structure in *Malus sieversii*, a wild progenitor species of domesticated apple. *Tree Genet. Genomes* 5:339–347.
- Sun, J., J.G. Fang, F. Wang, Q.B. Sun, and Z. Zhang. 2010. Characterisation of RNaseH-LTR sections of Ty1-copia retrotransposons in apple and fingerprinting of four apple clones by S-SAP analysis. *J. Hort. Sci. Biotechnol.* 85:53–58.
- Taberlet, P. and G. Luikart. 1999. Non-invasive genetic sampling and individual identification. *Biol. J. Linn. Soc. Lond.* 68:41–55.
- Tessier, C., J. David, P. This, J.M. Boursiquot, and A. Charrier. 1999. Optimization of the choice of molecular markers for varietal identification in *Vitis vinifera* L. *Theor. Appl. Genet.* 98:171–177.
- Urrestarazu, J., C. Miranda, L. Santesteban, and J. Royo. 2012. Genetic diversity and structure of local apple cultivars from northeastern Spain assessed by microsatellite markers. *Tree Genet. Genomes* in press.
- U.S. Department of Agriculture. 2012. Germplasm Resources Information Network (GRIN). 1 May 2012. <<http://www.ars-grin.gov/cgi-bin/npgs/html/index.pl>>.
- van Treuren, R., H. Kemp, G. Ernsting, B. Jongejans, H. Houtman, and L. Visser. 2010. Microsatellite genotyping of apple (*Malus × domestica* Borkh.) genetic resources in the Netherlands: Application in collection management and variety identification. *Genet. Resources Crop Evol.* 57:853–865.
- Venturi, S., L. Dondini, P. Donini, and S. Sansavini. 2006. Retrotransposon characterisation and fingerprinting of apple clones by S-SAP markers. *Theor. Appl. Genet.* 112:440–444.
- Volk, G.M., C.M. Richards, A.D. Henk, A.A. Reilley, P.A. Reeves, P.L. Forsline, and H.S. Aldwinckle. 2009. Capturing the diversity of wild *Malus orientalis* from Georgia, Armenia, Russia, and Turkey. *J. Amer. Soc. Hort. Sci.* 134:453–459.
- Volk, G.M., C.M. Richards, A.A. Reilley, A.D. Henk, P.L. Forsline, and H.S. Aldwinckle. 2005. Ex situ conservation of vegetatively propagated species: Development of a seed-based core collection for *Malus sieversii*. *J. Amer. Soc. Hort. Sci.* 130:203–210.
- Volk, G.M., C.M. Richards, A.A. Reilley, A.D. Henk, P.A. Reeves, P.L. Forsline, and H.S. Aldwinckle. 2008. Genetic diversity and disease resistance of wild *Malus orientalis* from Turkey and southern Russia. *J. Amer. Soc. Hort. Sci.* 133:383–389.
- Waits, L.P., G. Luikart, and P. Taberlet. 2001. Estimating the probability of identity among genotypes in natural populations: Cautions and guidelines. *Mol. Ecol.* 10:249–256.
- Wertheim, S.J. 1998. Rootstock guide: Apple, pear, cherry, european plum. Proefstation voor de fruitteelt, Wilhelminadorp, The Netherlands.
- Wünsch, A. and J.I. Hormaza. 2002. Cultivar identification and genetic fingerprinting of temperate fruit tree species using DNA markers. *Euphytica* 125:59–67.