GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Assessment of apple core collections constructed using phenotypic and genotypic data

W.H. Yun[1], S.H. Ban[1], G.H. Kim[1], J.-H. Kim[2], S.I. Kwon[2] and C. Choi[1]

[1]College of Agriculture and Life Science,
Kyungpook National University, Daegu, Korea
[2]Apple Experiment Station,
National Institute of Horticultural and Herbal Science,
Rural Development Administration, Gunwi, Korea

Corresponding author: C. Choi
E-mail: cc31@knu.ac.kr

Genet. Mol. Res. 14 (2): 6453-6464 (2015)
Received September 3, 2014
Accepted February 5, 2015
Published June 11, 2015
DOI http://dx.doi.org/10.4238/2015.June.11.21

ABSTRACT. Several types of information can be used to select core collections, including passport data, agronomic data, and molecular data. However, little is known about the ability of core collections to retain the genetic diversity and structure of the whole collection for characters that were not considered during the selection, particularly when molecular markers are used. In this study, two core subsets were established for the apple (*Malus* spp) germplasm bank curated at the Apple Research Station, National Institute of Horticultural and Herbal Science, Korea, based upon genetic diversity estimated with 14 simple sequence repeat markers, and phenotypic diversity based on 23 traits. Comparisons between these two subsets and with the whole collection were used to determine the effect of the data used in the selection on phenotypic and genetic diversity, and population structure. The two subsets had a similar diversity and did not differ from the original collection, according to the Nei and Shannon diversity indices. Allele and class frequencies were also maintained in the two subsets. Overall, the type of data used to construct the core collection had little influence

on the phenotypic and genetic diversity retained. Therefore, in the case of apple collections, the use of molecular markers is preferable, because they allow rapid and reliable characterization.

Key words: *Malus domestica*; Core collection; Cluster analysis; Simple sequence repeat; Germplasm

INTRODUCTION

The apple (*Malus domestica* Borkh.) is very important in Korea. In 2013, the worldwide production of apples was 76.38 tons, from about 4.84 million ha (FAO, 2015). Korean apple production in 2013 was 493,701 tons, which was 19.5% of the total fruit production (2.52 million tons). The apple-growing area in the country was 30,449 ha in size, which accounted for 19% of the total area (KOSIS, 2014).

Germplasm collections are important for crop improvement and research. Many countries and organizations have founded hundreds of gene banks, and millions of crop resources have been preserved (Tanksley and McCouch, 1997). However, with the continuous collection of germplasm resources, collections have grown larger and larger; this hinders the preservation, evaluation, and use of these resources. In order to utilize and manage germplasm collections more effectively and easily, Frankel (1984) proposed the "core collection" concept. A core collection is defined as a representative sample of the entire collection of a crop species and its relatives, with minimum repetitiveness and maximum genetic diversity. With a core collection, it is more convenient to study and utilize germplasm resources. Core collections of many plant species have been constructed, including *Medicago* spp (Basigalup et al., 1995), sesame (Xiurong et al., 2000), wheat (Balfourier et al., 2007), rice (Yan et al., 2007), peanut (Dwivedi et al., 2008), sorghum (Upadhyaya et al., 2009), and soybean (Oliveira et al., 2010).

Various types of data have been used to analyze the genetic diversity in collections, including morphological and ecogeographical traits, and molecular and biochemical markers (van Hintum et al., 2000), and each of these has its advantages and disadvantages. In large collections with thousands of accessions, such as many crop collections, assessing the entire collection with molecular markers is generally not feasible because of the cost (Grenier et al., 2000). Therefore, core collections for crop species have traditionally been obtained using a reduced set of ecogeographical and morphological data (Balakrishnan et al., 2000; Hu et al., 2000; Li et al., 2004). However, many fruit tree species have a much higher number of described phenotypic characters (50-70), and more than half of them are required for assessing distinctness (UPOV, 2005). Moreover, the population structure of conserved germplasm of most cultivated apple trees differs from that of annual species. Furthermore, field analyses are a lengthy process, as it may take several years after tree planting to obtain the first crop.

Currently, apple germplasm in Korea includes 178 wild species, 60 rootstocks, and 1080 improved varieties, and approximately 1340 additional species and varieties are to be collected. However, the utilization of genetic resources is very poor. Therefore, the effective collection, management, and breeding of the nation's apple genetic resources, and an effective way to build a core collection, are needed.

The main goal of this study was to determine the ability of morphological and molecular markers to construct core collections representative of the genetic and phenotypic diversity conserved in an apple collection maintained *ex situ*.

MATERIAL AND METHODS

Plant materials

The plant material used in this study consisted of 70 apple accessions maintained at the Apple Research Station, National Institute of Horticultural and Herbal Science, Rural Development Administration, Korea (Table 1).

Table 1. List of the 70 apple (*Malus* spp) cultivars included in the study.

No.	Accession	No.	Accession
1	*M. arnoldiana* "Sarg"	36	*M. hybrid* "Dolgo"
2	*M. asiatica*	37	*M. hybrid* "Eley Purple"
3	*M. zumi* "Nagasaki"	38	*M. hybrid* "France"
4	*M. arnoldiana* "Arnold Crab"	39	*M. hybrid* "Golden Hornet"
5	*M. asiatica* "Segeumjug"	40	*M. hybrid* "Hopa A"
6	*M. baccata* "Siberian"	41	*M. hybrid* "Hopa B"
7	*M. coronaria* "SI-12-70"	42	*M. hybrid* "Indian Magic"
8	*M. domestica* "Schlect Spur Red Delicious"	43	*M. hybrid* "Indian Summer"
9	*M. domestica* "Anoka"	44	*M. hybrid* "John Downie"
10	*M. domestica* "Binkwa"	45	*M. hybrid* "Okanagan"
11	*M. domestica* "Budagovsky 57-491"	46	*M. hybrid* "Ottawa 8"
12	*M. domestica* "Cheongdo"	47	*M. hybrid* "Profusion"
13	*M. domestica* "Darth Manta"	48	*M. hybrid* "Robinson"
14	*M. domestica* "Harcourt"	49	*M. hybrid* "Sentinel"
15	*M. domestica* "Hibernal"	50	*M. hybrid* "Shaguo"
16	*M. domestica* "Humboldt"	51	*M. hybrid* "SPY227"
17	*M. domestica* "Hwangsakbinkwa"	52	*M. hybrid* "Transcendent"
18	*M. domestica* "Mantet-1"	53	*M. hybrid* "Van Eseltine"
19	*M. domestica* "Mantet-2"	54	*M. hybrid* "Waka"
20	*M. domestica* "Matsumotonishiki"	55	*M. hybrid* "Whitney No. 10"
21	*M. domestica* "Meran M.78"	56	*M. hybrid* "Yantai"
22	*M. domestica* "Miyama"	57	*M. kansuensis* "Youngdonghadang"
23	*M. domestica* "Morihofu 3A Fuji"	58	*M. prunifolia* "Asami"
24	*M. domestica* "Ottawa 271"	59	*M. prunifolia* "Maruba"
25	*M. domestica* "Ottawa 274"	60	*M. robusta* "Robusta 5"
26	*M. domestica* "Sandongbinkwa"	61	*M. robusta* "Baily"
27	*M. domestica* "Suhongsakbinkwa"	62	*M. robusta* "Electa 88075"
28	*M. domestica* "Virginia"	63	*M. robusta* "Korea"
29	*M. domestica* "Virginiak-6"	64	*M. spectabilis* "Dadong 1"
30	*M. domestica* "Worcester Pearmain"	65	*M. spectabilis* "Flontish"
31	*M. domestica* "Zhigulevskoe"	66	*M. spectabilis* "Gorgeous"
32	*M. floribunda* "Hillieri"	67	*M. spectabilis* "Honghadang"
33	*M. hupehensis* "Hobookhadang"	68	*M. spectabilis* "Kwansang 1"
34	*M. hybrid* "Adams"	69	*M. spectabilis* "Michurina Tefseltone"
35	*M. hybrid* "Beverly"	70	*M. spectabilis* "Tartan"

Phenotypic analysis

The 70 apple cultivars were evaluated for 23 morphological characters according to UPOV Test Guidelines (Table 2). Standardized data and the PRINCOMP procedure for principal components analysis (PCA) were used to analyze genetic relationships between the apple cultivars, and select the most discriminant traits (Pereira-Lorenzo et al., 2003). Ten traits were selected from the 23 initial traits. Quantitative and qualitative traits were combined to perform a single cluster analysis, after the transformation of quantitative traits into qualitative ones. The number of classes (CN) for each quantitative trait was defined as follows (Santesteban et al., 2009) (Table 3):

$$CN = \frac{R_B}{\left(\frac{\sum_{i=1}^{n} R_{Ai}}{n}\right) + 1.15\left(\frac{\sum_{i=1}^{n} SD_{Ai}}{n}\right)} \quad \text{(Equation 1)}$$

where n is the number of accessions in the germplasm bank, R_B is the difference between the highest and lowest average trait values observed in the bank, R_A is the difference between the highest and lowest trait values observed within each accession, and SD_A is the standard deviation of the variable values observed within each accession. The *CN* value was rounded to the closest integer. The purpose of this procedure was to define class intervals that were broader than the average difference that could be found for each character within an accession, because the procedures available in the literature (Kaufman and Rosseeuw, 1990; Pecetti et al., 1992) were considered unsatisfactory.

Table 2. List of 23 horticultural traits and their range of values.

No.	Horticultural trait	Unit	Range
1	Leaf length	mm	56.9-120.9
2	Leaf width	mm	37.1-75.1
3	Leaf width/Leaf length	-	1.3-2.2
4	Petiole length	mm	16.0-41.2
5	Stipule length	mm	3.7-21.9
6	Stipule width	mm	0.5-7.1
7	Stipule width/Stipule length	-	2.3-15.5
8	Fruit weight	g	1.8-338.4
9	Fruit width	mm	11.8-80.5
10	Fruit length	mm	13.7-93.0
11	Fruit width/Fruit length	-	0.7-1.5
12	Calyx basin diameter	mm	0.3-16.4
13	Calyx basin length	mm	5.3-39.1
14	Fruit stalk diameter	mm	0.6-4.7
15	Fruit stalk length	mm	4.0-57.0
16	Stalk cavity diameter	mm	0.4-19.9
17	Stalk cavity length	mm	4.7-46.3
18	Fruit firmness	g/8 mm Ø	1.5-13.3
19	Fruit soluble solids content	ºBrix	9.0-27.9
20	Fruit acid	pH	0.3-3.1
21	L	-	26.9-75.2
22	a	-	-16.9-38.3
23	b	-	8.1-33.5

DNA extraction

Young leaves were collected from shoot tips of the 70 apple cultivars at the Apple Research Station and stored at -80°C until use. DNA was extracted using a Gentra Puregene Cell Kit (Qiagen), according to a modified protocol for plants.

Data analysis

Polymerase chain reaction (PCR) amplification was performed for 14 simple sequence repeat (SSR) markers (Table 4). The reaction mixture consisted of 20 ng of template DNA, 1 µM of each primer, and 1X Hot Start Taq Master Mix (PhileKorea Inc., Korea) in a

final volume of 10 µL. The PCR conditions were 5 min of pre-denaturation at 94°C, 35 cycles at 94°C for 30 s, 60°C for 30 s, and 72°C for 1 min, and 7 min of elongation at 72°C. The electrophoresis of the PCR products was performed using a Fragment Analyzer (Advanced Analytical Technologies Inc., USA).

Table 3. List of the quantitative phenotypic traits analyzed in the study.

Quantitative trait	Max	Min	R_B	Means R_A	Means SD_A	CN
Leaf length (mm)	120.9	56.9	63.99	22.65	18.34	2
Leaf width (mm)	75.1	37.1	38.06	17.45	26.65	2
Stipule length (mm)	21.9	3.7	18.16	7.04	4.39	2
Fruit weight (g)	338.4	1.8	336.65	31.48	25.94	5
Fruit width (mm)	80.5	11.8	68.75	7.58	4.51	5
Fruit length (mm)	93.0	13.7	79.22	8.55	23.42	2
Calyx basin diameter (mm)	16.4	0.3	16.09	4.79	6.35	2
Calyx basin length (mm)	39.1	5.3	33.79	5.45	7.96	2
Stalk cavity diameter (mm)	19.9	0.4	19.57	5.20	6.68	2

For each trait, the highest (Max) and lowest (Min) average trait values observed in the bank, their difference (R_B), the mean of the differences between the highest and lowest trait values observed within each accession (mean R_A), the mean of the standard deviation of the variable values observed within each accession (SD_A), and the number of qualitative classes defined (CN) are shown.

Table 4. Fourteen simple sequence repeat primers used in the study.

Marker	Forward/Reverse sequence (5'-3')	Allele size (bp)
Hi02C07	F: AGAGCTACGGGGATCCAAAT R: GTTTAAGCATCCCGATTGAAAGG	108-149
Chr02-0004	F: AGTGGAGTTTGGAAAGCCATC R: TGGTAATTTAGATTCCCGCAAG	242
GD12	F: TTGAGGTGTTTCTCCCATTGGA R: CTAACGAAGCCGCCATTTCTTT	141-191
Hi23g02	F: TTTTCCAGGATATACTACCCTTCC R: GTTTCTTCGAGGTCAGGGTTTG	230-257
CH04e03	F: TTGAAGATGTTTGGCTGTGC R: TGCATGTCTGTCTCCTCCAT	179-222
CH03d07	F: CAAATCAATGCAAAACTGTCA R: GGCTTCTGGCCATGATTTTA	186-226
Chr07-0048	F: TTCCATTAGTATCGCCTCAAGG R: ATGGGCTTCCCTTTTCTACAG	188
CH01f03b	F: GAGAAGCAAATGCAAAACCC R: CTCCCCGGCTCCTATTCTAC	139-183
Chr10-0018	F: ACAGTCGAGTGGCTCTTCTG R: TCTATGGTCGGACAGGTACGA	206
CH02d08	F: TCCAAAATGGCGTACCTCTC R: GCAGACACTCACTCACTATCTCTC	210-254
CH01F02	F: ACCACATTAGAGCAGTTGAGG R: CTGGTTTGTTTTCCTCCAGC	168-222
Chr13-0020	F: CGGAGGTGATACTCGCGTAA R: AAATCGTACCTGCAAAGCTGT	174
Chr14-0031	F: GCTTCGCAGTTTCGTGTACAAA R: GGGAATACTCATTTCCTCGAAT	158
Chr17-0014	F: CGTTTGCAGAATTTGGAACTG R: AGATGTGGTGAACGTGTCAAC	250

Construction of core collections

The procedure of stepwise clustering with random sampling proposed by Hu et al. (2000) was used to develop two core subsets, one from each source of data: the CG subset was

obtained from microsatellite data, and the CP subset was obtained from morphoagronomic traits. The simple matching coefficient method and the unweighted pair group method with arithmetic mean (UPGMA) were used for hierarchical cluster analysis (Sneath and Soksl, 1973) to group accessions. The stepwise procedure was as follows: (1) Genetic distances between accessions were calculated and accessions were classified by hierarchical cluster analysis based on these genetic distances; (2) One accession was randomly removed from a subgroup with the least distance, and another accession of the subgroup was sampled; (3) The genetic distances between the remaining accessions were calculated, and the sampling was repeated in the same way. The stepwise cluster analysis was conducted until the selected accessions were reduced to approximately 20% of the initial collection (Crossa et al., 1995; Yonezawa et al., 1995). The number of accessions eventually retained in each core collection was 14 (20%) for the CG subset and 14 (20%) for the CP subset. The analyses were conducted and phenograms were generated using NTSYSpc, ver. 2.11w (Rohlf, 1993).

Evaluation of the diversity in the collections

The genetic diversity index (DI) per locus and trait (corrected for sample size) was calculated for each core subset, and for the whole collection (WC), as follows:

$$DI = \frac{n(1 - \sum_{i=1}^{n} p_i)}{n - 1} \quad \text{(Equation 2)}$$

where n is the number of phenotypic classes or alleles and pi is the proportion of the total number of accessions in the ith class (Nei, 1987). Diversity was also estimated for each locus and trait using the Shannon diversity index (Shannon, 1948) as follows:

$$H' = -\sum_{i=1}^{n} p_i \log(p_i) \quad \text{(Equation 3)}$$

Because of the additive nature of the Shannon diversity index (Poole, 1974), values were pooled for each character. Differences in the Shannon diversity index between the subsets and the WC were determined by chi-square test. Frequency distributions of phenotypic traits and SSR markers were compared between the two subsets and the WC, using a chi-square goodness-of-fit test. The tests were performed only on the most frequent alleles and classes (those with a frequency higher than 10%) (Santesteban et al., 2009).

RESULTS

Phenotypic analysis

For the 70 apple varieties, 23 horticultural traits were investigated to determine the range of values for each trait. The ranges' minimum average values in the bank to maximum average values were as follows: length, 56.9-120.9 mm; leaf width, 37.1-75.1 mm; petiole length,

16.0-41.2 mm; stipule length, 3.7-21.9 mm; stipule width, 0.5-7.1 mm; fruit weight, 1.8-338.4 g; fruit width, 11.8-80.5 mm; fruit length, 13.7-93.0 mm; calyx basin diameter, 0.3-16.4 mm; calyx basin length, 5.3-39.1 mm; fruit stalk diameter, 0.6-4.7 mm; fruit stalk length, 4.0-57.0 mm; stalk cavity diameter, 0.4-19.9 mm; stalk cavity length, 4.7-46.3 mm; fruit firmness, 1.5-13.3 g/8 mm Ø; soluble solids content, 9.0-27.9 °Brix; and fruit acid, 0.3-3.1.

The horticultural traits examined had different units of measurement, and the variance was small. Variables with small variance receive less weight in PCA; therefore, PCA was performed after standardization. The eigenvalues and the contribution of each principal component are presented in Table 5. Of the 23 horticultural traits included in the analysis, the first principal component represented about five, and the second and third components represented about two to three. Together, the first three principal components explained 63.3% of the variation: 38.3% was explained by the first principal component, 13.8% by the second, and 11.2% by the third. An eigenvalue greater than 1 indicates that the principal component accounts for more variance than accounted for one of the original variables; the cumulative contribution of the first 12 principal components was 96.6%.

Table 5. Eigenvalues and contributions of the 23 principal components obtained using 23 horticultural traits in 70 apple cultivars.

Principal component	Eigenvalue	Contribution (%)	Cumulative contribution (%)
PC1	8.81	38.3	38.3
PC2	3.17	13.8	52.1
PC3	2.57	11.2	63.3
PC4	1.77	7.7	71.0
PC5	1.23	5.3	76.3
PC6	1.22	5.3	81.6
PC7	0.92	4.0	85.6
PC8	0.84	3.6	89.2
PC9	0.55	2.4	91.6
PC10	0.49	2.1	93.7
PC11	0.36	1.6	95.3
PC12	0.29	1.3	96.6
PC13	0.21	0.9	97.5
PC14	0.17	0.7	98.2
PC15	0.12	0.5	98.7
PC16	0.10	0.4	99.1
PC17	0.05	0.2	99.3
PC18	0.05	0.2	99.5
PC19	0.02	0.1	99.6
PC20	0.02	0.1	99.7
PC21	0.02	0.1	99.8
PC22	0.00	0.00	99.8
PC23	0.00	0.00	99.8

SSR marker analysis

The genetic diversity of the 70 apple cultivars was analyzed using 14 SSRs (Table 6). The total number of alleles was 501. The number of alleles per locus ranged from 21 for marker Chr10-0018 to 52 for marker CH03d07. The average number of alleles per locus was 36. The genetic diversity was the highest for Hi23g02 and CH03d07 (0.9693) and the lowest for Chr10-0018 (0.9169), and the average was 0.9510. The polymorphism information content (PIC) of the 14 SSR markers used in the analysis ranged between 0.9120 (Chr10-0018) and 0.9684 (Hi23g02 and CH03d07), and the mean PIC was 0.9486.

Table 6. Number of alleles, genetic diversity, and polymorphism information content (PIC) of 14 simple sequence repeats analyzed in 70 apple cultivars.

Locus	No. of alleles	Genetic diversity	PIC
Hi02C07	32	0.9341	0.9304
Chr02-0004	34	0.9574	0.9558
GD12	25	0.9348	0.9310
Hi23g02	47	0.9693	0.9684
CH04e03	46	0.9637	0.9624
CH03d07	52	0.9693	0.9684
Chr07-0048	22	0.9370	0.9335
CH01f03b	35	0.9407	0.9377
Chr10-0018	21	0.9169	0.9110
CH02d08	34	0.9526	0.9505
CH01F02	45	0.9684	0.9674
Chr13-0020	35	0.9572	0.9555
Chr14-0031	32	0.9483	0.9459
Chr17-0014	43	0.9637	0.9624

Comparison of SSRs in the WC and the core subsets

A total of 501 alleles were recorded for the 14 analyzed SSR loci, 91 of which were present in more than 10% of the accessions (frequent alleles). In the WC, the mean allelic richness was 35.8 alleles per locus, and the number of alleles per locus ranged from 21 at locus Chr10-0018 to 52 at locus CH03d07. The gene diversity was relatively high, and ranged between 0.92 and 0.97 (mean of 0.95) (Table 7). A comparison of the WC and the core subsets revealed similar tendencies: no matter which data were originally used to create the subset, the mean allelic richness was of the same magnitude in the two subsets (18.8 and 18.4 alleles per locus for the CG and CP subsets), and all of the frequent alleles were present in the WC and the two subsets. The retained gene diversity was high; ranges and mean values in both the CG subset and the CP subset were similar to those in the WC. The Shannon diversity index values (H') for the WC and the two subsets were also similar (H' = 3.18 ± 0.51 for the WC, and 2.45 ± 0.28 and 2.46 ± 0.29 for the CG and CP subsets, respectively). When compared with the WC, the two core subsets had similar allele distributions for all 14 loci (Table 8). There were no significant differences between the allele distributions of the CG and CP subsets.

Table 7. Simple sequence repeat diversity assessed by the number of frequent alleles (A) in the whole collection (WC), the number of rare alleles (B) in the WC, the number of alleles that were rare in the WC and present in each subset (C), and the genetic diversity (DI) as defined by Nei (1987).

Locus	WC			CG		CP	
	A	B	DI	C	DI	C	DI
Hi02C07	6	26	0.93	11	0.90	12	0.93
Chr02-0004	3	31	0.96	15	0.92	15	0.94
GD12	10	14	0.93	7	0.93	6	0.92
Hi23g02	3	44	0.97	21	0.95	23	0.96
CH04e03	6	39	0.96	14	0.92	12	0.93
CH03d07	6	46	0.97	19	0.95	15	0.93
Chr07-0048	9	13	0.94	7	0.92	7	0.89
CH01f03b	6	29	0.94	11	0.93	11	0.91
Chr10-0018	10	11	0.92	5	0.91	4	0.89
CH02d08	7	27	0.95	10	0.93	13	0.94
CH01F02	4	41	0.97	18	0.95	18	0.94
Chr13-0020	6	29	0.96	15	0.94	9	0.91
Chr14-0031	9	23	0.95	8	0.93	11	0.92
Chr17-0014	6	37	0.96	11	0.94	10	0.93
Total	91	410		172		166	
Mean	6.5	29.3	0.95	12.3	0.93	11.9	0.93

CG = core subset created from microsatellite data; CP = core subset created from morphological data.

Table 8. Comparison of the frequency distribution of simple sequence repeat markers in the whole collection (WC), in the two subsets, and between the two subsets, using a χ^2 test.

Locus	Comparison with the whole collection				Comparison between subsets	
	CG *vs* WC		CP *vs* WC		CG *vs* CP	
Hi02C07	0.06	NS	0.03	NS	0.01	NS
Chr02-0004	0.15	NS	0.19	NS	0.00	NS
GD12	0.02	NS	0.03	NS	0.00	NS
Hi23g02	0.61	NS	0.61	NS	0.00	NS
CH04e03	0.23	NS	0.20	NS	0.00	NS
CH03d07	0.52	NS	0.32	NS	0.04	NS
Chr07-0048	0.03	NS	0.08	NS	0.01	NS
CH01f03b	0.06	NS	0.08	NS	0.00	NS
Chr10-0018	0.01	NS	0.03	NS	0.01	NS
CH02d08	0.08	NS	0.16	NS	0.01	NS
CH01F02	0.38	NS	0.40	NS	0.00	NS
Chr13-0020	0.25	NS	0.18	NS	0.01	NS
Chr14-0031	0.07	NS	0.07	NS	0.00	NS
Chr17-0014	0.38	NS	0.19	NS	0.05	NS

CG = core subset created from microsatellite data; CP = core subset created from morphological data; NS = non significant difference at the $P < 0.05$ level.

In the WC, 410 alleles had a frequency lower than 10%. Of these rare alleles, 172 were maintained in the CG subset and 166 were maintained in the CP subset. The rare alleles lost in the two subsets were always present in fewer than ten accessions of the WC.

Comparison of the core subsets and the WC

The mean phenotypic richness in the WC was 2.90 classes per trait, and the phenotypic diversity ranged from 0.30 for stipule length to 0.76 for fruit (mean of 0.48) (Table 9). As with the molecular markers, comparisons between the WC and the CP subset showed that all of the frequent classes (25) were retained, and the diversity index values were similar. Finally, the Shannon diversity index values were not significantly different between the WC and the two subsets ($H' = 1.00 \pm 0.52$ for the WC, and 0.85 ± 0.44 and 1.01 ± 0.48 for the CP and CG subsets, respectively).

Table 9. Morphological diversity assessed by the number of frequent classes (A) in the whole collection (WC), the number of rare classes in the WC (B), the number of classes that were rare in the WC and were present in each subset (C), and the genetic diversity (DI), as defined by Nei (1987).

Locus	WC			CG		CP	
	A	B	DI	C	DI	C	DI
LL	2	0	0.49	0	0.41	0	0.49
LW	2	0	0.49	0	0.46	0	0.49
SL	2	0	0.30	0	0.34	0	0.46
FW	1	4	0.49	3	0.53	3	0.37
FL	5	0	0.76	0	0.76	0	0.69
FWL	2	0	0.43	0	0.46	0	0.34
CBD	2	0	0.43	0	0.46	0	0.34
CBL	2	0	0.41	0	0.41	0	0.24
FSD	2	0	0.43	0	0.46	0	0.41
FSL	3	0	0.54	0	0.58	0	0.54
Total	25	4		3		3	
Mean	2.5	0.40	0.48	0.30	0.49	0.30	0.44

CG = core subset created from microsatellite data; CP = core subset created from morphological data; LL = leaf length; LW = leaf width; SL= stipule length; FW = fruit width; FL = fruit length; FWL = fruit width/length ratio; CBD = calyx basin diameter; CBL = calyx basin length; FSD = fruit stalk diameter; FSL = fruit stalk length.

There were no significant differences between the class distributions of the CG and CP subsets (Table 10). Of the four rare classes observed in the WC, two were maintained in the CP subset and three were maintained in the CG subset. Once again, the rare classes lost in the core subsets were present in fewer than ten accessions of the WC.

Table 10. Comparison of the frequency distribution of phenotypic traits between the whole collection (WC), the two subsets, and between the two subsets, using a χ^2 test.

Trait	Comparison with the whole collection				Comparison between subsets	
	CG *vs* WC		CP *vs* WC		CG *vs* CP	
LL	0.00	NS	0.01	NS	0.01	NS
LW	0.00	NS	0.00	NS	0.00	NS
SL	0.06	NS	0.00	NS	0.03	NS
FW	0.07	NS	0.00	NS	0.07	NS
FL	0.04	NS	0.00	NS	0.04	NS
FWL	0.02	NS	0.00	NS	0.03	NS
CBD	0.02	NS	0.00	NS	0.03	NS
CBL	0.06	NS	0.00	NS	0.09	NS
FSD	0.00	NS	0.00	NS	0.00	NS
FSL	0.00	NS	0.01	NS	0.01	NS

CG = core subset created from microsatellite data; CP = core subset created from morphological data; LL = leaf length; LW = leaf width; SL = stipule length; FW = fruit width; FL = fruit length; FWL = fruit width/length ratio; CBD = calyx basin diameter; CBL = calyx basin length; FSD = fruit stalk diameter; FSL = fruit stalk length; NS = non significant difference at the $P < 0.05$ level.

DISCUSSION

The two core subsets developed in this study were found to be representative of the phenotypic and genetic diversity of the collection. They retained all of the frequent alleles or classes present in the collection and all of the rare alleles or classes present in more than ten accessions in the collection. No losses of diversity or alterations in frequency distributions were observed. Since the concept of the core collection was proposed (Brown, 1989), various types of data, such as morphological, agronomical, and ecogeographical traits, or molecular and biochemical markers, have been used to analyze genetic diversity (van Hintum et al., 2000). To our knowledge, the ability of these different types of data to core collections, while adequately retaining the overall diversity within a collection, has been analyzed only by Grenier et al. (2000) in sorghum in a limited manner. Overall, our results confirm the preliminary study of Grenier et al. (2000), and suggest that phenotypic and genetic diversity are not affected by the type of character used to construct the core collection. Therefore, this particular issue seems to be of little relevance when deciding what type of data should be used, so other criteria, such as time, cost, or ease of data acquisition should be taken into account. In large collections with thousands of accessions, such as many crop collections, assessing the entire collection with molecular markers is generally not feasible because of the cost (Grenier et al., 2000). Therefore, core collections for crop species have traditionally been obtained using a reduced set of passport, ecogeographical, and morphological data (Balakrishnan et al., 2000; Hu et al., 2000; Li et al., 2004), and only at a later stage has the genetic diversity of the core collection been assessed with molecular markers (Grenier et al., 2000). However, there are a sufficient number of reasons to take a different approach to constructing core collections for many species. Although these collections usually consist of fewer than a thousand accessions

(Maggioni et al., 2004), many fruit tree species have a much higher number of described phenotypic characters (50-70), and more than half of them are required for assessing distinctness (UPOV, 2005). Moreover, the population structure of the conserved germplasm of most cultivated apple trees differs from that of annual species, because in vegetatively propagated fruit tree species the domestication process usually involves few recombination cycles, and, consequently, domesticated genotypes are only a few generations apart from their wild ancestors (Escribano et al., 2008). Therefore, the morphological characterization of apple collections can involve more cost and effort than assessing the collections with molecular markers (Karp et al., 1997). Additionally, Hu et al. (2000) found that genetic sorting based on phenotypic data does not correctly reflect the genetic diversity of the initial germplasm resources, owing to errors in the field or genotype-environment interactions. Therefore, in regional and national apple collections not yet characterized, molecular markers could constitute a rapid and economic tool for the estimation of genetic variability and the construction of core subsets, so that in a later stage it will be feasible to perform an exhaustive and efficient characterization of the core collection using agromorphological descriptors.

ACKNOWLEDGMENTS

Research supported by a grant from the Next-Generation BioGreen ‘‘21’’ Program (#PJ01105602), Rural Development Administration, Korea.

REFERENCES

Balakrishnan R, Nair NV and Sreenivasan NV (2000). A method for establishing a core collection of *Saccharum officinarum* L. germplasm based on quantitative morphological data. *Genet. Resour. Crop Evol.* 47: 1-9.

Balfourier F, Roussel V, Strelchenko P, Exbrayat-Vinson F, et al. (2007). A worldwide bread wheat core collection arrayed in a 384-well plate. *Theor. Appl. Genet.* 114: 1265-1275.

Basigalup DH, Barnes DK and Stucker RE (1995). Development of a core collection for perennial *Medicago* plant introductions. *Crop Sci.* 35: 1163-1168.

Brown HD (1989). Core collections: a practical approach to genetic resources management. *Genome* 31: 818-824.

Crossa J, Delacy IH and Taba S (1995). The use of multivariate methods in developing a core collection. In: Core collections of plant genetic resources (Hodgkin T, Brown AHD, van Hintum THJL, Morales EAV, eds.). John Wiley Sons, New York, 77-89.

Dwivedi SL, Puppala N, Upadhyaya HD, Manivannan N, et al. (2008). Developing a core collection of peanut specific to Valencia market type. *Crop Sci.* 48: 625-632.

Escribano P, Viruel MA and Hormsza JJ (2008). Comparison of different methods to construct a core germplasm collection in woody perennial species with simple sequence repeat markers. A case study in cherimoya (*Annona cherimola*, Annonaceae), an underutilized subtropical fruit tree species. *Ann. Appl. Biol.* 153: 25-32.

FAO (Food and Agricultural Organization) (2015). FAOSTATA. FAO, Korea.

Frankel OH (1984). Genetic perspectives of germplasm conservation. In: Genetic manipulation: impact on man and society (Arber W, Illmensee K, Peacock WJ and Starlinger P, eds.). Cambridge University Press, Cambridge, 161-170.

Grenier C, Deu M, Kresovich S, Bramel-Cox PJ, et al. (2000). Assessment of genetic diversity in three subsets constituted from the ICRISAT sorghum collection using random *vs* non-random sampling procedures. B. Using molecular markers. *Theor. Appl. Genet.* 101: 197-202.

Hu J, Zhu J and Xu HM (2000). Methods of constructing core collections by stepwise clustering with three sampling strategies based on the genotypic values of crops. *Theor. Appl. Genet.* 101: 264-268.

Karp A, Kresovich S, Bhat KV, Ayad WG, et al. (1997). Molecular tools in plant genetic resources conservation: a guide to the technologies. *IPGRI Tech. Bull*. 2: 47.

Kaufman L and Rosseeuw PJ (1990). Finding groups in data: An introduction to cluster analysis. Wiley, New York, 342.

KOSIS (Korean Statistical Information Service) (2014). Statistical Database, KOSIS, Korea.

Li Y, Shi Y, Cao Y and Wang T (2004). Establishment of a core collection for maize germplasm preserved in Chinese National Genebank using geographic distribution and characterization data. *Genet. Resour. Crop Evol.* 51: 845-852.

Maggioni L, Fischer M, Lateur M, Lamont EJ, et al. (2004). Report of a working group on Malus/Pyrus, Second meeting, 2-4 May 2002, Dresden-Pillnitz, Germany. IPGRI, Rome, 96.

Nei M (1987). Molecular evolutionary genetics. Columbia University Press, New York, 512.

Oliveira MF, Nelson RL, Geraldi IO, Cruz CD, et al. (2010). Establishing a soybean germplasm core collection. *Field Crops Res.* 119: 277-289.

Pecetti L, Annicchalarico P and Damania AB (1992). Biodiversity in a germplasm collection of durum wheat. *Euphytica* 60: 229-238.

Pereira-Lorenzo S, Ramos-Cabrer AM, Ascasibar-Errasti J and Pineiro-Andion J (2003). Analysis of apple germplasm in northwestern Spain. *J. Am. Soc. Hortic. Sci.* 128: 67-84.

Poole RW (1974). An introduction to quantitative ecology. McGraw-Hill, New York, 532.

Rohlf FJ (1993). NTSYS-PC. Numerical taxonomy and multivariate analysis system. Exeter Software, New York.

Santesteban LG, Miranda C and Royo RB (2009). Assessment of the genetic and phenotypic diversity maintained in apple core collections constructed by using either agro-morphologic or molecular marker data. *Span. J. Agric. Res.* 7: 572-584.

Shannon CE (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* 27: 379-423

Sneath PHA and Soksl RR (1973). Numerical taxonomy. Freeman, San Francisco, 573.

Tanksley SD and McCouch SR (1997). Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* 277: 1063-1066.

Upadhyaya HD, Pundir RPS, Dwivedi SL, Gowda CLL, et al. (2009). Developing a mini core collection of sorghum for diversified utilization of germplasm. *Crop Sci.* 49: 1769-1780.

UPOV (2005). Guidelines for the conduct of tests for distinctness, homogeneity and stability (Apple), TG/14/9. International Union for the Protection of New Varieties of Plants, Geneva, Switzerland, 45.

van Hintum THJL, Brown AHD, Spillane C and Hodgkin T (2000). Core collections of plant genetic resources. *IPGRI Tech. Bull.* 3: 48.

Xiurong Z, Yingzhong Z, Yong C, Xiangyun F, et al. (2000). Establishment of sesame germplasm core collection in China. *Genet. Resour. Crop Evol.* 47: 273-279.

Yan W, Rutger JN, Bryant RJ, Bockelman HE et al. (2007). Development and evaluation of a core subset of the USDA rice germplasm collection. *Crop Sci.* 47: 869-878.

Yonezawa K, Nomura T and Morishima H (1995). Sampling strategies for use in stratified germplasm collections. In: Core collections of plant genetic resources (Hodgkin T, Brown AHD, van Hintum THJL and Morales EAV, eds.). John Wiley Sons, New York, 35-53.