African Journal of Biotechnology Vol. 9 (2), pp. 123-127, 11 January 2010
Available online at http://www.academicjournals.org/AJB
ISSN 1684–5315 © 2010 Academic Journals

Full Length Research Paper

Sampling strategy to develop a primary core collection of apple cultivars based on fruit traits

Zhang Jie[1], Wang Yi[1], Zhang Xinzhong[1], Li Tianzhong[1], Wang Kun[2], Xu Xuefeng[1] and Han Zhenhai*

[1]Institute of Horticultural Plants, China Agricultural University, a Key Laboratory of Beijing Municipality of Stress Physiology and Molecular Biology for Fruit, Tree, Beijing 100193, China.
[2]Fruit Tree Research Institute of the Chinese Academy of Agricultural Sciences, China.

Accepted 8 May, 2009

A total of 435 accessions of apple germplasm collected from the Xincheng National Apple Germplasm Repository and 10 morphological traits of them were used for studying the optimal sampling strategy for primary core collection of apple (*Malus domestica* Brokh). In order to acquire the appropriate primary core collection, different entire sampling ratio and sampling scheme were compared in the study. Six entire sampling ratios were tested and the sampling schemes following stratification into two levels, including the grouping principle, sampling proportion within group were studied. The results showed that 10% should be the suitable entire sampling ratio for primary core collection of apple. Under 15% entire sampling ratio, the optimal sampling scheme was grouped based on cultivar group combining with genetic diversity based sampling proportion within group proportion. This sampling strategy was used to acquire the primary core collection of 64 accessions from 435 accessions of apple cultivars, and the primary core collection could well represent the genetic diversities of the entire variety collection.

Key words: Core collection, sampling strategy, fruit traits, apple.

INTRODUCTION

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, 1997). However, with continuous collection of germplasm resources, the sizes of collections have been becoming larger and larger; this hinders the preservation, the evaluation, the research and the use of germplasm resources. In order to utilize and manage the germplasm collection more effectively and easily, Frankel et al. (1984) proposed the 'Core Collection' concept (Frankel et al., 1984). A core collection is defined as a representative sample of the entire collection with minimum repetitiveness and maximum genetic diversity of a crop species and its relatives. With the core collection, it is convenient to study and utilize germplasm resources, which has been received by more and more researchers in the entire world. Up to the present, the core collections of many plant were constructed in the world, which include rice, wild bean, capsicum, midicago, peanut, barley, peach, strawberry and so on (Zichao, 2002; Tohme, et al., 1996; Zewide et al., 2004; Diwan et al., 1995; Corley et al., 2005; Yin-xia et al., 2007; Geibel et al., 2004).

Apple, as one of the four fruits in the world, is of great importance in China. As one of apple centre of origin, abundant apple germplasm resources exist in China. Moreover, with continuous collection and selection of apple resources, more than 700 accessions of apple germplasm had been preserved in China. The huge amount disturbed the study and utilization of apple germplasm. Therefore, developing an apple core collection is apparently essential for the most efficient conservation, management and utilization of the whole germplasm collections.

One common approach for constructing a core collection is grouping the entire collection according to growing regions or ecotypes, then, selecting representative core accessions from each of the groups to form core subsets,

*Corresponding author. E-mail: rschan@cau.edu.cn.

and the entire core collection is constructed by combining all core subsets. Therefore accessions need to be analyzed on genetic diversity to ensure their representativeness (Wang et al., 2007). For a core collection, the representativeness is the most important property. The representativeness evaluation is a significant step in the procedure of core collection construction. A series of evaluating parameters is required in representativeness evaluation for core accessions. Therefore, selecting a series of evaluating parameters that can make impact is an important aspect of core collection construction research. The key point to improve the representativeness of a core collection is the reliable stratification on the initial collection and scientific section within groups. Only if the representativeness of core subsets from groups is improved, the representativeness of the entire core collection could be substantially improved. A core subset is a watch of a core collection, and its evaluating parameters are almost the same as in core collection.

MATERIALS AND METHODS

Plant materials

The data used in this study were provided by the National Apple Database of China. The plant material includes 437 accessions collected from 25 provinces that cover six ecological areas in the country. Among the 18 descriptors recorded in the database, 12 were complete for all accessions. These 10 descriptors include original region, fruit form, fruit texture, fruit color, fruit juice, fruit flavor, soluble matter content, single fruit weight, generational growth period and maturation date.

Sampling strategies and statistical analysis

According to their ecological distribution and accession origin (Li, 2001), the apple accessions were divided into 3 groups. Five commonly used sampling strategies, namely random, constant, proportional, logarithmic and genetic diversity-based were compared to select core accessions from each group based on a fixed sampling size (Ortiz et al.,1998; Zichao, 2002; Yin-xia, 2006). The five strategies are briefly described as follows.

In random strategy (R), entries are randomly sampled from the entire collection, ignoring groups; in proportion strategy (P), entries are sampled in each group in proportion to the group size. The limitation of this strategy is the under-representation in groups containing few samples but with large genetic diversity. Square root strategy (S) entails entries being sampled in proportion to the square root of group size; logarithmic strategy (L) involves entries being sampled in proportion to the logarithm of group size. Square root and Logarithmic strategies tend to enlarge or reduce the sampling ratio in the small groups and the large groups, respectively. While in genetic diversity-based strategy (G), entries are sampled in proportion to the amount of genetic diversity in the groups. Genetic diversity-based strategy considers both proportion and genetic diversity of groups.

Genetic diversity index and coefficient of variance were calculated by the following formulae:

Genetic diversity index I is

$$I = -\frac{\sum_i \sum_j P_{ij} \log P_{ij}}{N}$$

Where P_{ij} is the frequency of some trait, i is the ith type and j is the jth type in a section.

The reservation proportion of phenotype is

$$RPR = \frac{\sum_i M_i}{\sum_i M_{i0}}$$

Where M_i is the number of the ith type in the established primary core collection. M_{i0} is the number of the ith type in the entire core collection. i is the ith type in a section.

The maximum range Dmax is

$$D\max = \frac{STD_i(Max_i - Max_{i0})}{N}$$

Where STD_i is standardized treatment of the ith type in the section, Max_i is the maximum of the ith type in the established primary core collection; Max_{i0} is that of the entire core collection. N is the total number of types in the section.

The minimum range Dmin is

$$D\min = \frac{STD_i(Min_i - Min_{i0})}{N}$$

Where STD_i is standardized treatment of the ith type in the section, Min_i is the minimum of the ith type in the established primary core collection; Min_{i0} is that of the entire core collection. N is the total number of types in the section.

Variation quantity CV is

$$CV = \frac{\frac{\sum_j (X_{ij} - \bar{X}_i)^2}{\sum_i \frac{Mi-1}{\bar{X}_i}}}{N}$$

Where X_{ij} is the value of the ith traits in the jth materials; $\bar{X}_i$ stands for the mean of the ith traits; Mi stands for the number of traits in a section; N stands for the total number of types in the section.

The phenotype variance is

$$VPV = \frac{STD_i\left[\sum_i \frac{\sum_j (X_{ij} - \bar{X}_i)^2}{M_i - 1}\right]}{N}$$

The STD_i is standardized treatment of the ith type in the section; X_{ij} is the value of the ith traits in the jth materials; $\bar{X}_i$ stands for the mean of the ith traits; M_i stands for the number of traits in the primary core collection; N stands for the total number of types in the section.

Construction of the core collection

Accessions were selected to constitute the core collection according to their origins, geographical stratification, sampling strategy, cluster and random sampling. The number of accessions chosen for the core collection depends on the size of the entire collection and the sampling ratio. For a given sampling strategy, an appropriate sampling ratio to constitute the core collection was

Table 1. Rank analysis by test parameters and comparison of ratio of phenotype retained under different origin samples ratios.

Sampling percentage	index of genetic diversity	variance of phenotype value	Average rank	Ratio of phenotype retained
10%	1.0	1.0	1.00	93.1%
20%	1.0	1.5	1.25	95.7%
30%	1.5	2.0	1.75	97.5%
40%	2.0	2.0	2.00	99.1%
50%	2.0	3.0	2.50	99.9%
60%	2.5	3.0	2.75	100%

determined by taking into account the coincidence between the sampled collection and the original collection, geographic distribution and altitude of the original accessions.

Evaluation of the primary core collections

The coincidence between the entire collection and the established primary core collection was analyzed in the aspects of trait coverage, character variation pattern, genetic diversity and geographical distribution pattern. The maximum, minimum, range, variation coefficient, phenotypic variance and reservation proportion of fruit growth period, single fruit weight and soluble solid matter content were used as evaluating parameters. The quantity trait reservation proportion is the trait range of established primary core collection that divides the entire core collection, and the range is the trait maximum that subtracts its minimum.

RESULTS

Total sampling ratio of the core subset to the initial germplasm population

Six sampling ratio of the core subset to the initial germplasm population were designed in this study, which were from 10 to 60%. Table 1 showed that when the total sampling ratio was 10%, the value of average rank was 1.0 and the ratio of phenotype retained was up to 93.1%, indicating that the genetic redundancy of primary core collection was lowest but its genetic variance abundance was higher. Therefore, 10% was most suitable sampling ratio for developing apple primary core collection.

Grouping principles and sampling proportion within group

Core collections were suggested to improve germplasm utilization. A core collection is a subset chosen to represent the diversity of a collection with a minimum of redundancies. Because diversity is distributed between and within groups with different degrees of organization, an adequate classification of accessions into related groups should be performed prior to the selection of a core collection. In the study, six test parameters of three different grouping principles and four most commonly used sampling proportion were compared for the suitability in selecting apple core accessions (Table 2). According to the average rank of three grouping principles, the rank value of cultivar grouping was the minimum, followed by maturation date grouping, and the value of origin region was the maximum. This demonstrated that the effect of cultivar grouping was the best, followed by maturation date grouping, and the effect of origin region was the worst.

For the sampling proportion within group, there was significant difference only at rank value of *Dmin* and *Dmax* among different sampling proportion and no difference in other parameters. From the average rank of sampling proportion in group, we can see that the effect of genetic diversity proportion is the best. Thus, it is clear that genetic diversity proportion was most suitable sampling proportion in group.

Evaluation of primary core collection in apple

From Table 3, we can see that the reserved ratios of three quantitative traits were higher than 75%. Except for the soluble matter content, the *CV* of the primary collection of growth periods and single fruit weight content were higher than that of the whole collection, indicating that there were abundant variant types in the primary core collection.

The diversity of 7 traits for both the primary collection and the entire collection is listed in Table 4. From the data, there were higher coincidence between the primary core collection and the whole collection. The traits, including growth period, fruit juice and maturation date were completely retained. The coincidence of fruit shape was the lowest, but it was higher than 70%. Therefore, we can make the conclusion that almost all the 7 characters that existed in the entire collection also appeared in the primary core collection. Both qualitative and quantitative were similar between the entire collection and constructed core collection. The entire collection is well-represented by the core collection.

DISCUSSION

Sample size is another important issue to be addressed when establishing a crop core collection. The sampling

Table 2. Rank analysis by six test parameters of primary core collection constructed by different grouping principles and sampling proportion within group.

Principle	Group	Test parameters						
		I	*PRP*	*Dmin*	*Dmax*	*VPV*	*CV*	Average rank
Grouping principles	Cultivar group	1.0	1.0	1.0	1.0	1.0	1.0	1.00
	Maturation date	1.0	1.5	1.0	1.0	2.0	1.5	1.33
	Origin region	2.0	2.0	2.0	1.5	2.5	2.0	2.00
Sampling proportion within group	Simple proportion	1.0	1.0	1.0	1.5	1.0	1.0	1.08
	Square root proportion	1.0	1.0	1.5	1.5	1.0	1.0	1.17
	Logarithmic proportion	1.0	1.0	1.5	1.5	1.0	1.0	1.17
	Genetic diversity proportion	1.0	1.0	1.0	1.0	1.0	1.0	1.00

Table 3. Comparison of primary core collection and the total collection for quantitative traits

Traits	Primary collection				Whole collection				Reserved ratio (%)
	Max	Min	Range	CV (%)	Max	Min	Range	CV (%)	
Growth periods	247	172	75	10.4	247	165	82	9.2	91.5
Single fruit weight	350.1	40	310.1	38.1	350.1	30.1	320	31.8	96.9
Soluble matter content (%)	14.5	8.9	5.6	12.1	16.1	8.8	7.3	13.7	76.7

Table 4. Coincidence of traits between the core collection and the whole collection.

Traits	Primary collection	Whole collection	Coincidence (%)
Fruit texture	7	8	87.5
Fruit flavor	7	8	87.5
Fruit color	10	16	75
Fruit shape	10	12	83.3
Growth period	3	3	100
Fruit juice	4	4	100
Maturation date	5	5	100

size should depend upon genetic diversity, degree of genetic redundancy among the whole accessions. It is an important work to confirm reasonable sampling percentage in core collection construction. Brown suggested that core collection with 10% sampling percentage could represent 70% genetic diversity of the initial population when the number of the initial accessions was over 3000 (Brown, 1989). Zewdie (2004) used l0% sampling percentage in sorghum core collection construction. Li et al. (2007) also used l0% sampling percentage in primary core collection of peach cultivars. Xu et al. (2004) developed a cotton core collection with 30% sampling percentage (168 initial accessions). Wang et al. (2004) constructed a Changjiang soybean with just 8.58% sampling percentage. Zhao et al. (2005) established a core collection for the Chinese annual wild soybean with 9.8% sampling percentage.

Generally, small sampling percentage is suitable for large size initial population, while large one is suitable for small size initial population (Zichao et al., 2002; Chandra et al., 2002). In the study, we also used 10% sampling percentage to select the apple primary core collection. Compared with the average rank and ratio of phenotype retained by 10% sampling percentage and other sampling percentage, it is clear that 10% was the suitable sampling percentage.

The accessions for core collection were selected according to their origins, sampling strategy and random sampling. The number of accessions chosen for the core collection depends on the size of the entire collection and the sampling ratio. It has been reported that grouping prior to sampling is the key point for constructing a core collection (Hu et al., 2000; Li et al., 2004). The selection of entries in each group is the most important factor for the development of a core collection. A good sample of the core collection should maintain the maximum genetic diversity of the species in question. Random, constant, square root, logarithmic and genetic diversity-based are

the five commonly used ways for sampling. It has been proposed that a stratified strategy, weighted by either the

group size or genetic diversity, is better than a random or constant strategy.

The most important criterion for a good core collection is its representation for the entire collection, which can be evaluated by what degree the core collection maintains the full range of genetic variation of the entire collection. There are two types of evaluating methods: one is to compare the pre-core collection with the entire collection, and then compare it with the fixed core collection; the other is to compare the core collection with the entire collection directly (Diwan et al., 1994). In the present study, the 10 morphological characters, variation coefficient, genetic diversity indexes were compared between the core collection and the entire collection, both quailtative and quantitative were similar between the entire collection and constructed core collection. Therefore, the entire collection is well-represented by the core collection.

ACKNOWLEDGEMENTS

This study was supported by the Beijing Natural Science Foundational of China (6071002) and the National High Technology Research and Development Program ("863"Program) of China (2006AA10Z1B6). We thank the National Apple Germplasm Repository and Institute of fruit trees in Hebei province to provide apple germplasm information and materials.

REFERENCES

Brown AHD (1989). Core collections: A practical approach to genetic resources management, Genome, 31: 818-824.

Chandra S, Huaman Z, Krishna HS Ortiz R (2002). Optimal sampling strategy and core collection size of Andean tetraploid potato based on isozyme data—A simulation study, Theor. Appl. Genet. 104: 1325-1334.

Corley H, Dong W (2005). Development and Evaluation of a Mini Core Collection for the U.S. Peanut Germplasm Collection. Crop Sci. 45: 1540-1544.

Diwan N, Mcintosh Bauchan GR (1995). Methods of developing a core collection of annual Midicago species. Theor. Appl. Genet. 90(6): 755-761.

Frankel OH (1984a). Genetic perspectives of germplasm conservation, in Genetic Manipulation: Impact on Man and Society (eds. Arber W, Llimensee K, Peacock WJ Starlinger, P Cambridge: Cambridge University Press, pp. 179-183.

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

Geibel M, Masny A, Coman M (2004). The European strawberry database and building up a European core collection. Acta Hort., 649: 41-44.

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

Li CT, Shi CH, Wu JG, Xu HM, Zhang HZ, Ren YL (2004). Methods of developing core collections based on the predicted genotypic value of rice (*Oryza sativa* L.), Theor. Appl. Genet. 108: 1172-1176.

Li Y (2001). Researches of apple germplasm resources. (in Chinese)

Li YX, Gao QJ, Li TH (2007). Sampling strategy based on fruit characteristics for a primary core collection of peach cultivars. Euro. J. Horticult. Sci. 72(6): 268-274.

Li ZC, Zhang HL, Zeng YW，Yang ZY, Shen SQ, Sun CQ, Wang XK. (2002). Studies on sampling for establishment of core collection of rice landrace in Yunnan, China. Genet. Res. Crop Evol. 49(1): 67-72.

Ortiz R, Ruiz-Tapia EN, Mujica-Sanchez A (1998). Sampling strategy for a Core Collection of Peruvian quinoa germplasm, Theor. Appl. Genet. 90: 475-483.

Tanksley SD, McCouch SR（1997). Seedballkand molecular maps: Unlocking genetic Potential from the wild. Sci. 277: 1063-1066.

Tohme J, Gonzalez D (1996). AFLP analysis of gene pools of a wild bean core collection. Crop Sci. 36: 1375-1384.

Wang J, Hu J, Zhang CF, Zhang S (2007). Assessment on evaluating parameters of rice core collections constructed by genotypic values and molecular marker information. Rice Sci. 14(2): 101-110.

Wang LX, LI YH, Liw, Zhu L, Guan Y Ning XC, Guan RX, Liu ZX, Chang RZ, Qiu LJ (2004). Establishment of a core collection of Changjiang spring sowing soybean. Biodiv. Sci. 12: 578-585.

Xu HM, Qiu YX, Hu J, Wang JC (2004). Methods of constructing core collection of crop germplasm by comparing different genetic distances, cluster methods and sampling strategies. Acta Agron. Sin, 30: 932-936.

Zewdie Y, Tong NK, Bosland P (2004). Establishing a core collection of *capsicum* using a cluster analysis with enlightened selection of a ccessions. Genet. Res. Crop Evol., 51: 147-151.

Zhao L, Dong Y, Liu B, Hao S, Wang K, LI X (2005). Establishment of a Core Collection for the Chinese annual wild soybean. Chin. Sci. Bull. 10: 989-996.