academicJournals

Vol. 12(37), pp. 5577-5587, 11 September, 2013
DOI: 10.5897/AJB12.1263
ISSN 1684-5315 ©2013 Academic Journals
http://www.academicjournals.org/AJB

African Journal of Biotechnology

*Full Length Research Paper*

# Establishment of core collection from apricot germplasm in China

**Zhang Jun Huan, Wang Yu Zhu*, Sun Hao yuan, Yang Li and Ning Ning**

Institute of Forestry and Pomology, Beijing Academy of Agriculture and Forestry Science, Beijing, 100093, China.

Accepted 21 August, 2013

**This study aimed at establishing a core collection based on the analysis of data from simple sequence repeat (SSR) alleles and morphological and agronomical traits (MOR) of the primary core collection from apricot germplasm resources. The index of genetic diversity, and frequency ratios of retention and loss of the alleles were studied between cluster and random sampling methods at five sampling ratios. The results demonstrate that the cluster sampling method preceded random sampling, and cluster sampling of SSR combined with MOR at the rate of 80% was the best sampling strategy among all the sampling methods. Based on this sampling strategy, 120 accessions were selected as the core collection of apricot, which retained 100% alleles in the primary core collection and 100% phenotypic characters. The core collection developed had also been evaluated by using the data of six quantitative traits, which showed that the established core collection could well represent the genetic diversity of the original collection of 1501 apricot accessions.**

**Key words:** Apricot, core collection, sampling strategy, simple sequence repeat (SSR) molecular marker.

## INTRODUCTION

Apricot is an important fruit crop in the world. Usually, it is consumed fresh or dry, but processing is also favored due to their specific taste, aroma and nutritive values. As the origin center of apricots, China has the richest resources of both wild and cultivated varieties, including 2000 varieties or types (Wang, 1998), which provide considerable opportunities for genetic research and breeding. However, huge numbers of accessions represent challenges for their conservation, evaluation, identification and utilization (Grenier et al., 2000; Tanksley et al., 1997).

Moreover, for perennial woody plants, large plantation area and high cost of manage-ment places a severe restriction on their size. So it is more urgent to establish core collection of Chinese apricot resources. The concept of core collection was initiated by Frankel in 1984, defined as a representative sample of the whole collection with minimum repeti-tiveness of the 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. In apricot, we firstly performed the construction of primary core collection from 1501 accession of apricot in China using mor-phological data (Wang et al., 2011). The size of the primary core collection is still so large that redundancy of some accessions may occur because apricot is usually clonally propagated. Therefore, it is necessary to develop a core collection with the same genetic diversity as the whole collection, but smaller in size than the primary core

*Corresponding author. E-mail: chinabjwyz@sohu.com. Tel: +86 10 82592521. Fax: +86 10 62598744.

**Abbreviations:** ***Ia*,** Index of genetic diversity of alleles; ***Ip*,** index of genetic diversity of phenotype; **MOR,** morphologic and agronomic traits; **RRa,** retained ratio of alleles; **RRp,** retained ratio of phenotype.

collection.

To characterize and evaluate effectively apricot germplasm, molecular markers such as random amplified polymorphic DNA (RAPD), amplified fragement length polymorphism (AFLP), simple sequence repeat (SSR) and inter simple sequence repeat (ISSR) were developed. Among these markers, SSRs are highly polymorphic, informative, co-dominant, technically simple and reproducible, and become common in developing a core collection for some crops (Yao et al., 2008; Zhang et al., 2009a). In addition, when to construct the core collection of a crop, the morphologic data are usually applied extensively because of those data recorded relative comprehensively (Li et al., 2002, 2007; Zhang et al., 2010). However, there were few reports about developing a crop collection using morphologic data combining with SSR data.

The objective of this study was to develop a core collection based on the analysis of Simple Sequence Repeat (SSR) alleles and morphological data (MOR) of the primary core collection of apricot in China, and so as to provide valuable references to the scientific conservation and utilization.

## MATERIALS AND METHODS

The primary core collection of apricot, comprising 150 accessions, was used in this study. This collection was sampled from the initial collection including 1501 accessions of apricot in China and represented 100% of the diversity in the initial collection at the morphological level (Wang et al., 2011).

### Morphological data

Eighteen (18) morphological and agronomical traits were selected in this study. They consist of (1) fruit size, (2) fruit shape, (3) fruit height, (4) lateral width, (5) ventral width, (6) symmetry in ventral view, (7) suture, (8) depth of stalk cavity, (9) shape of apex, (10) ground color, (11) hue of over color, (12) color of flesh, (13) texture of flesh, (14) juice content, (15) soluble solid content, (16) adherence of stone to flesh, (17) ripe stage and (18) flesh fiber content, including 6 quantitative variables and 12 qualitative class variables, which were all related to fruit morphology. These phenotypic characters of about 150 accessions apricot germplasm, collected from the apricot repository of Beijing and Xiongyue, Liaoning province, were recorded for two years using the 57 defined descriptors for apricot by International Union for the Protection of New Varieties of Plants (UPOV) (TG /70/4, 2005) and the book ‘descriptors and data standard for apricot (*Prunus armeniaca* Mill)’ for observation on fruit fruit (Liu and Liu, 2006), 25 typical fruits were selected from each of the analyzed accession. The material was grafted on apricot seeding rootstocks with at least three replications per genotype.

According to the method of Li et al. (2002), the quantitative characters were quantified into 10 categories, where the distance between two neighboring categories was every 0.5 standard deviation.

### SSR data

Total genomic DNA was extracted from fresh leaf following the method of Doyle and Doyle (1987). PCR amplification and electrophoresis were performed as described by Wang et al. (2011). The SSR primers (Table 1) were designed based on the nucleotide sequences reported by Testolin et al. (2000), Sosinski et al. (2000), Dirlewanger et al. (2002), Lopes et al. (2002) and Hagen et al. (2004). These 22 pairs of primers were firstly screened for amplification of polymorphic and unambiguous bands in the 150 accessions. An example of amplification products of SSR primer pair (UDP97-401) is shown in Figure 1. Only clear, well defined and reproducible bands were recorded for developing a core collection. Presence (1) and absence (0) of each band was scored. The dataset was converted into a mathematical matrix, which was used to perform statistical analysis and calculate the number of alleles by using Cervus version 2.0 and PopGene version 1.32 softwares. The polymorphic microsatellite markers identified in this study is listed in Table 1. A total of 196 alleles were identified at 22 SSR loci in 150 accessions from the apricot primary core collection. The number of alleles per locus ranged from 5 to 15, with an average of 8.91 alleles per locus.

In the 22 loci, the expected heterozygosity values varied greatly, with an average of 0.731. In addition, the average value of polymorphic information content (PIC) was 0.695, ranging from 0.44 to 0.86.

### Sampling scheme of apricot core collection

A flowchart of the methodology used in the establishment of core collection is presented in Figure 2. Sampling schemes were developed at two levels, that is, the sampling method and sampling proportion. The sampling methods included cluster analysis based on 3 different data (SSR, MOR, SSR + MOR) and random sampling. For the size of core subset under each sampling method, sampling proportion from the primary collection was designed as 50, 60, 70, 80, 90 and 100%. Combining the sampling methods with the sampling proportion, 24 sampling strategies were used to develop core collection, and thus, 24 core collections were generated. In order to avoid losing some important biological types, the sample selecting of the core collection were conducted according to the determined sampling strategies in combination with many other germplasm information at the same time. Some accessions with distinct traits will be deliberately added to the list of the core collection if not being selected.

### Evaluating parameters for sampling methods

Three evaluating parameters were selected according to the study of Li et al. (2002, 2007) with some modification, including index of genetic diversity (I) of phenotype (or alleles), retained ratio (RR) of phenotype (or alleles), frequency and ratio of loss of alleles. Formulas for *I* and *RR* are as follows:

$$I = \frac{-\sum_{i}\sum_{j} P_{ij} \ln P_{ij}}{N}, \quad RR = \frac{\sum_{i} M_{i}}{\sum_{i} M_{io}}$$

Where, $P_{ij}$ is the frequency of the $j^{th}$ phenotype (or alleles) in the $i^{th}$ trait (or locus); $N$ is the total number of traits (loci); where, $M_{i0}$ is the number of the $i^{th}$ phenotype (or alleles) of the initial germplasm group; $M_i$ is the number of the $i^{th}$ phenotype (or alleles) of core subset.

The loss frequency of alleles ($P$) was divided into two ladders, that is $P \leq 0.01$ and $0.01 < P \leq 0.03$; the loss ratio of alleles at different $P$ range = the lost number of alleles in one core subset / the lost

**Table 1.** Polymorphic microsatellite markers identified in the construction of apricot core collection.

| Locus | Ogrin | Repeat motif | No. of alleles | PIC | Heterozygosity |
|---|---|---|---|---|---|
| UDP96-005 | *P. persica* | $(AC)_{16}TG(CT)_2CA(CT)_{11}$ | 9 | 0.648 | 0.694 |
| UDP97-401 | *P. persica* | $(GA)_{19}$ | 10 | 0.638 | 0.680 |
| UDP97-402 | *P. persica* | $(AG)_{17}$ | 7 | 0.700 | 0.741 |
| UDP98-406 | *P. persica* | $(AG)_{15}$ | 11 | 0.848 | 0.866 |
| UDP98-409 | *P. persica* | $(AG)_{19}$ | 9 | 0.768 | 0.800 |
| UDP98-411 | *P. persica* | $(TC)_{16}$ | 11 | 0.810 | 0.834 |
| UDP98-412 | *P. persica* | $(AG)_{28}$ | 9 | 0.829 | 0.849 |
| Pchcms4 | *P. persica* | $(CA)_9$ | 6 | 0.440 | 0.470 |
| Pchgms4 | *P. persica* | $(CT)_{21}$ | 7 | 0.549 | 0.624 |
| Pchgms10 | *P. persica* | $T_{19}A_{10}$ | 6 | 0.480 | 0.562 |
| BPPCT001 | *P. persica* | $(GA)_{27}$ | 6 | 0.567 | 0.612 |
| BPPCT002 | *P. persica* | $(AG)_{25}$ | 8 | 0.797 | 0.825 |
| BPPCT012 | *P. persica* | $(CT)_{13}CC(CT)_7$ | 11 | 0.684 | 0.706 |
| BPPCT028 | *P. persica* | $(TC)_{15}$ | 5 | 0.525 | 0.606 |
| BPPCT029 | *P. persica* | $(GA)_{12}(CAGA)_4$ | 10 | 0.776 | 0.800 |
| BPPCT030 | *P. persica* | $(AG)_{25}$ | 5 | 0.476 | 0.512 |
| ssrPaCITA15 | *P. armeniaca* | $(TC)_{15}$ | 10 | 0.787 | 0.814 |
| ssrPaCITA19 | *P. armeniaca* | $(TC)_{16}$ | 15 | 0.860 | 0.876 |
| AMPA095 | *P. armeniaca* | $(AC)_{13}T(AC)_4$ | 13 | 0.674 | 0.694 |
| AMPA105 | *P. armeniaca* | $(AG)_{11}$ | 11 | 0.843 | 0.862 |
| AMPA109 | *P. armeniaca* | $(TG)_{11}$ $(AG)_9$ | 8 | 0.780 | 0.810 |
| AMPA112 | *P. armeniaca* | $(AG)_{18}$ | 9 | 0.814 | 0.838 |
| Average | | | 8.91 | 0.695 | 0.731 |

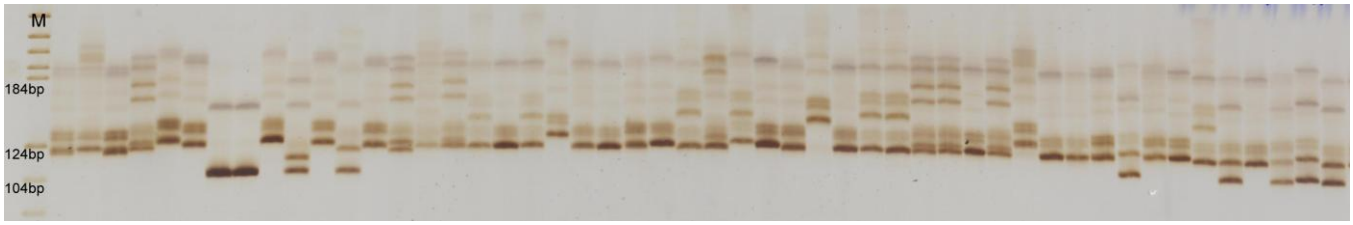


**Figure 1.** An example of SSR primer pair (UDP97-401) amplification patterns in some apricot accessions (M: Marker).

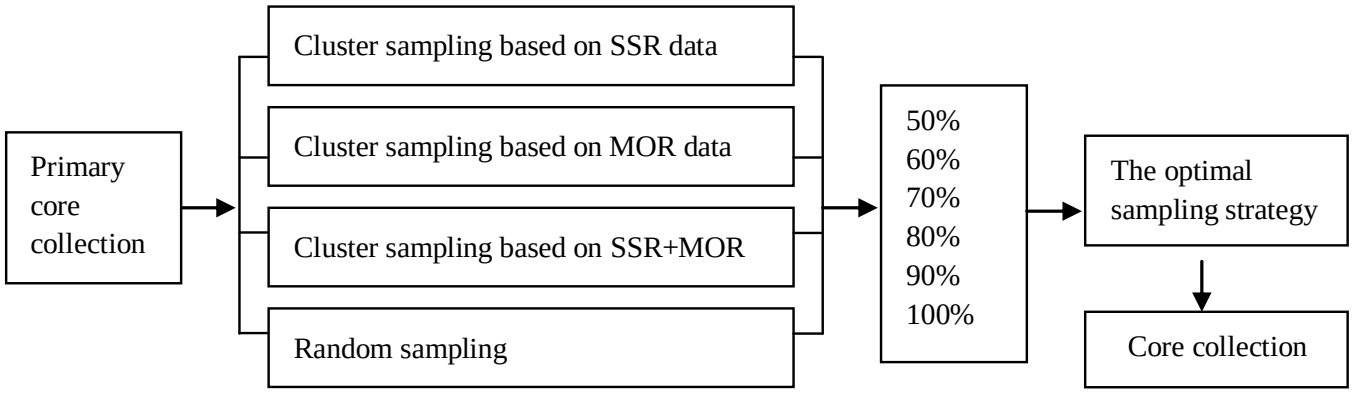


**Figure 2.** Method for establishment for the core collection of apricot cultivars.

number of alleles in the primary core collection, the *R* value of one trait in the initial collection.

**Evaluation of the representation of core collection based on the validated sampling methods**

To determine the representation of the core collection, 6 quantitative traits and 6 evaluating parameters were selected to compare the entire and core collection. The 6 traits include fruit weight, fruit height, lateral width, ventral width, soluble solid content and ripening time. The 6 evaluating parameters are maximum, minimum, range *(R)*, coefficient of variation (*CV*), variance of phenotype value (*VPV*) and ratio of trait retained (*RTR*). The values of *R*, *CV*, *VPV* and *RTR* were calculated using the following formulas, respectively.

*R* = the maximum value of one trait — the minimum value of one trait:

$$CV = \frac{\sum \frac{\sqrt{\sum \left(X_{ij} - \overline{X}_i\right)^2 / n-1}}{\overline{X}_i}}{N}$$

Where, $X_{ij}$ is the phenotypic value of the $j^{th}$ accession in the $i^{th}$ trait, $\overline{X}_i$ is the mean of phenotypic values of all accessions in the $i^{th}$ trait, *n* is the number of accessions, *N* is the total number of traits.

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

Where, $STD_i$ is the standardization for the $i^{th}$ trait; $X_{ij}$ is the phenotypic value of the $j^{th}$ accession in the $i^{th}$ trait; $\overline{X}_i$ is the mean of phenotypic values of all accessions in the $i^{th}$ trait, $M_i$ is the number of the $i^{th}$ phenotype in core collection, *N* is the total number of traits.

$$RTR\ (\%) = \frac{\text{The } R \text{ value of one trait in the core subset}}{\text{The } R \text{ value of one trait in the initial collection}} \times 100$$

The values of *I* was calculated according to the aforementioned formulas.

## RESULTS

### The analysis of the index of genetic diversity among four sampling methods at different sampling proportion

For the index of genetic diversity of phenotype (*Ip*), as shown in Figure 3a, when randomly sampling methods were used, there was no regularity in the changes of the *Ip* values among different sampling proportion. When cluster sampling was based on SSR + MOR, the *Ip* value reached the most at the sampling percentage of 50%, and then slightly decreased with the sampling percentage increased. Whereas, for other two cluster sampling methods, the *Ip* value presented an increasing trend with the sampling ratio increased. At the same sampling ratio of 60, 70 and 80%, the *Ip* values in the core collection using cluster sampling methods based on MOR was higher than that using other two cluster sampling methods. It can be seen from Figure 3b, for the index of genetic diversity of alleles (*Ia*), it was also no regularity in the changes of the *Ia* values when randomly sampling method was used to construct the core collection. At any sampling ratio point, the *Ia* value of the core collection using cluster sampling methods based on MOR was all lower than that using other two cluster sampling methods. Whereas, the highest level occurred in the cluster sampling methods based on SSR + MOR. For the cluster sampling methods based on SSR, the *Ia* values of alleles were increased with the sampling ratio increased, and reached the highest at the sampling percentage of 80%.

Comparing Figure 3a with 3b, it can be seen that the value of the index of genetic diversity in allele was obviously higher than that in phenotype at the same sampling ratio. Under most sampling ratios, the *Ip* values were the highest when using cluster sampling methods based on MOR, followed by SSR + MOR, the lowest was SSR. However, for the *Ia* value, it reached the highest when using cluster sampling methods based on SSR + MOR. The value of the index of genetic diversity for one core collection is higher, represented the core collection is more perfect. All together, the cluster sampling methods based on SSR + MOR was more preferred than the other three sampling methods.

### The analysis of retained ratios among four sampling methods

The retained ratios of phenotype (*RRp*) were analyzed and compared among the 24 core collections established according to different sampling strategies (Figure 4a). Similar to the index of genetic diversity, when randomly sampling methods were used, there were no regular changes in *RRp* values among different sampling proportion. For the three cluster sampling methods, the *RRp* values of the core collections rapidly increased with the sampling percentage increased. At any sampling percentage, the *RRp* values of the core collection constructed using cluster sampling based on MOR was maximum among the four sampling methods, and at sampling percentage of 60%, the core collection was able to preserve 100% phenotype of the primary collection. Comparing MOR + SSR with SSR, the former methods was relatively suitable; the *RRp* values reached higher to

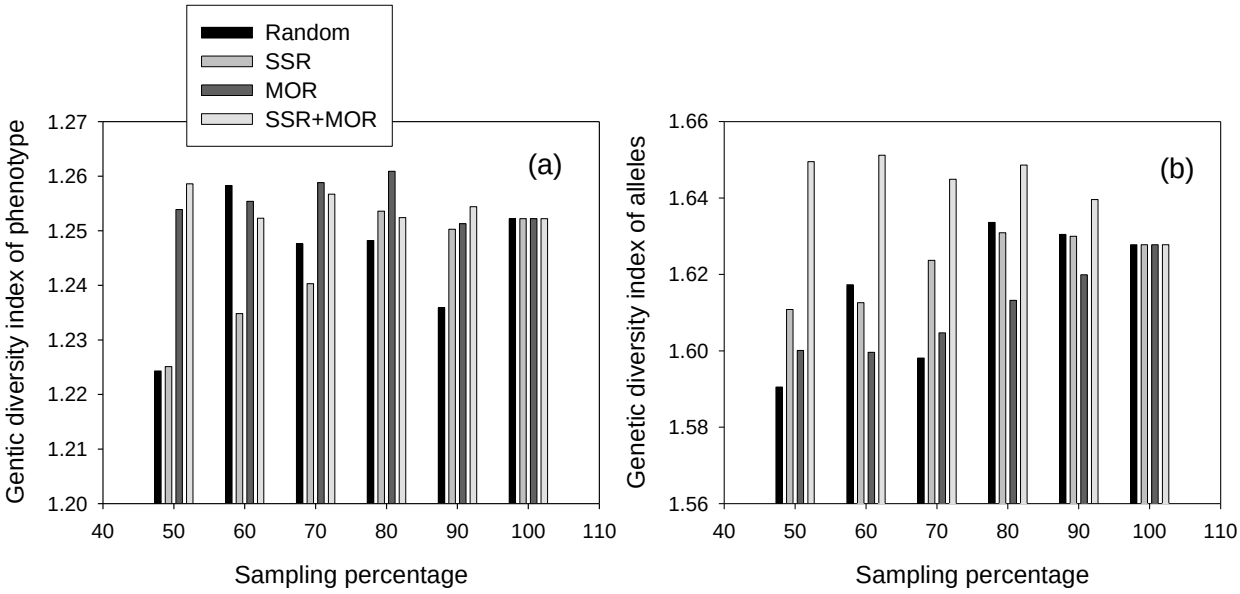


**Figure 3.** Influence of different sampling methods on genetic diversity index of core collections.

98.39% at sampling percentage of 50%. Whereas, at the same sampling ratio, the *RRp* values was only 96.77% when using cluster sampling based on SSR. A total of 22 pairs of SSR primers were selected for amplification of high polymorphic and unambiguous bands. A total of 196 alleles were identified at 22 SSR loci in 150 accessions from the apricot primary core collection. Statistically analyzed the number and the frequency of the alleles of the core collections using different sampling strategies, and calculated the retained ratios of alleles (*RRa*). As presented in Figure 4b, for the four sampling methods, the *RRa* values of the core collections almost increased with the sampling percentage increased. The *RRa* values of the core collection constructed using cluster sampling based on MOR + SSR was obviously higher than that using other three sampling methods at corresponding sampling percentage.

Comparing Figure 4a with 4b, it can be seen that the *RRp* values was higher than the *RRa* values at the same sampling ratio among the four sampling methods. For example, at sampling percentage of 50%, the *RRp* values were higher than 96%; whereas, the *RRa* values were only 92%. In addition, under most sampling ratios, the *RRp* and *RRa* values of the core collections constructed using cluster sampling based on MOR + SSR all remained higher than the other three sampling methods.

## Analysis of the loss of alleles among four sampling methods

The objective of the development of core collection is to select a representative sample of the whole collection with minimum repetitiveness and maximum genetic diversity of a crop species and its relatives. During the process of constructing the core collection, it is inevitable that the loss of alleles will occur when the genetic repetitiveness was eliminated. The lower frequency allele was usually easy being lost during the sampling compressed. The loss ratio of alleles at different frequency range among different sampling strategies was analyzed in the present study. The result is shown in Figure 5. At similar sampling percentage, the loss ratio of alleles was highest when $P \leq 0.01$ among the four sampling methods. Where, the $P$ is the frequency of a lost allele presented in the primary core collection. Comparatively, the loss ratio of alleles was much lower when $0.01<P \leq 0.03$, and all the alleles of the primary core collection would remained in the core collection when $0.03< P \leq 0.06$.

As shown in Figure 5a, the alleles of $P \leq 0.01$ were lost in various degrees among the four sampling methods. At the sampling percentage from 50 to 70%, for the three cluster sampling methods, the loss ratio of alleles decreased as the sampling percentage increased. The loss ratio of alleles in the core collection constructed using cluster sampling based on MOR was the highest than that using other three sampling methods. According to this method, the loss ratio of alleles was still higher than 0.27 even if the sampling percentage increased higher to 80%. Correspondingly, the loss ratio of alleles was all lower than 0.17 for the other three sampling methods.

The loss ratio of alleles was much lower of $0.01<P \leq 0.03$ than that of $P \leq 0.01$. As the sampling percentage increased, the loss ratios of alleles were all rapidly decreased among all the sampling methods

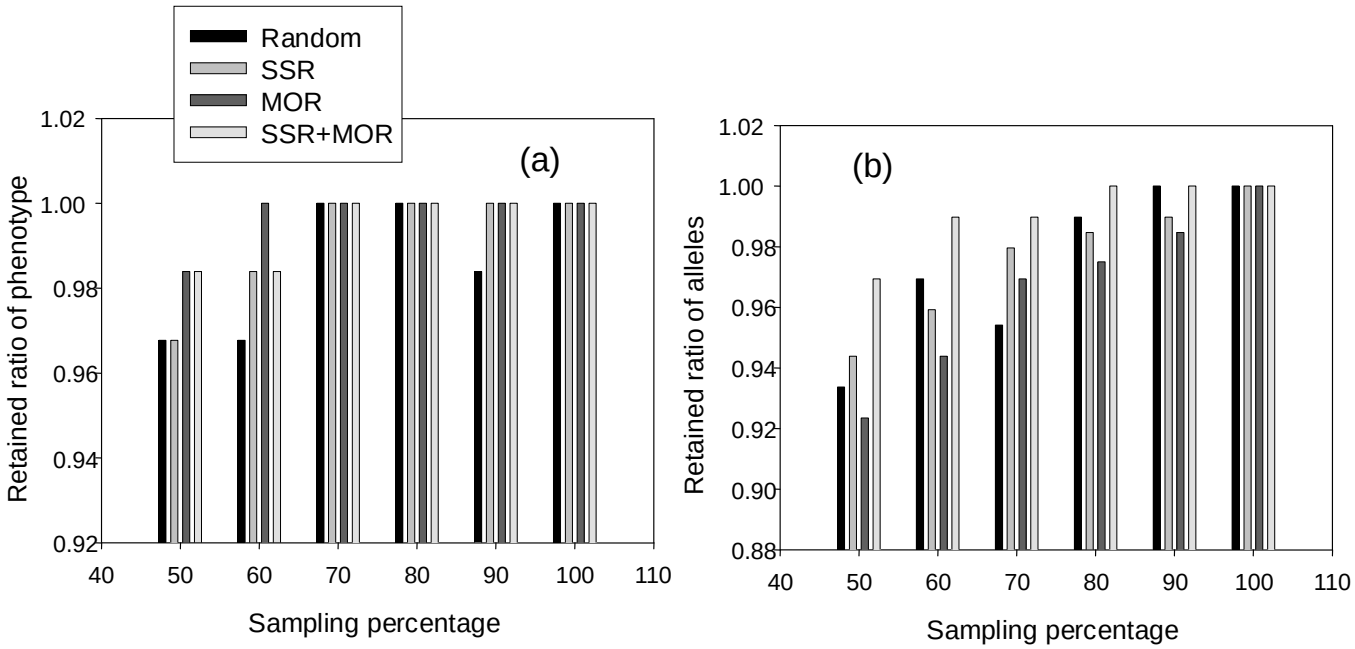


**Figure 4.** Influence of different sampling methods on retained ratio of core collections.

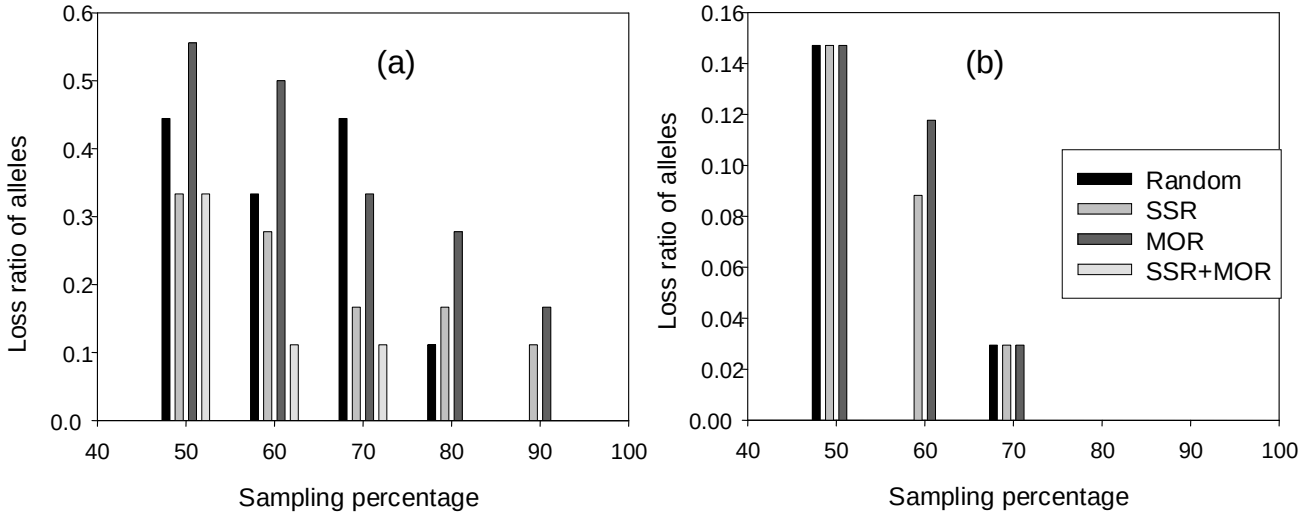


**Figure 5.** Analysis of ratio of alleles lost under different sampling methods. (a) $P \leq 0.01$ (b) $0.01 < P \leq 0.03$.

(Figure 5b).

These results indicate that it was an effective way to retain these alleles of $0.01 < P \leq 0.03$ by increasing the sampling percentage. At the same sampling percentage, the loss ratio of alleles of the core collection constructed using cluster sampling based on SSR + MOR was the most low, and the value was decreased to zero at sampling percentage of 50%. When sampling percentage increased to 80%, the loss ratio values were also decreased to zero. For the alleles of $0.03 < P \leq 0.06$, the loss ratio values were all zero among all the sampling strategies, that is, all the alleles of $0.03 < P \leq 0.06$ all remained in the core collection constructed using any sampling methods. Above these results, it was difficult to retain the alleles of the lower *P* value, even if the sampling percentage increased. However, for the alleles of the higher *P* value, the loss ratios would be rapidly decreased with the sampling percentage increased. Thus, it is an effectively way to retain these alleles of the higher *P* value by increasing the sampling percentage.

### Determination of the apricot core collection

According to the analysis of the 24 core collections established based on 24 sampling strategies, it is suggested that the optimal sampling methods were cluster sampling based on SSR + MOR, and the suitable sampling percentage was 80%. Using this sampling strategy, we have established the apricot core collection with 120 accessions, including 75 accessions for flesh market and processing, 11 accessions for kernel consuming, 7 accessions for ornamental cultivars, 7 accessions for newly bred cultivars (or lines), 13 accessions for introduced germplasm and 7 accessions for inter-specific hybrids. The name of germplasm and their characteristics are listed in Table 2.

The core collection remained both 100% alleles of the primary core collection and all the phenotype traits of the initial collections, and the valuable germplasm with different economic traits (for example, fruit size, fruit quality, early ripeness, self fertility, disease tolerance, cold resistance, etc.) were all remained in the core collection. For example, ‘Liquanerzhuanzi’ is one apricot cultivar of large fruit size, the soluble solids content of fruit in ‘Keziximixi’ is high to 21% and its Vc value was also high to 22.1 mg/l00 g FW, ‘Anjiana’ is a cultivar of high sugar and the content was high to 9.9%, ‘80A03’ is a genotype with very large kernel and the average weight of kernel is high to 0.98, ‘Luotuohuang’ is an early ripening cultivar (fruit development period is 55 days), ‘Kaite’ (originally name is Katy) is a cultivar with high ratio of self-fertility (Cao et al., 2012), and the blossom characteristics is special such as ‘Liaomeixing’ with multiplied petal and ‘Lveshanxing’ with green sepal, and so on.

### Estimation of the apricot core collection

Five parameters of 6 quantitative traits from the core collection were compared with those of same parameters from the primary and initial collection, and the results are listed in Table 3. Compared with the primary core collection, except soluble solid content, the values of the three parameters (maximum, minimum and range) of other five quantitative traits in the core collection were all same as that in the primary core collection. The *CV* and *VPV* values of five traits (except fruit ventral width) were higher in the core collection than those in primary core collection, indicating that the core collection eliminated some genetic redundancy and can represent genetic diversity of the primary core collection very well.

In comparison with initial collection, the *CV* and *VPV* values of soluble solid content were lower in the core collection than those in initial collection; whereas, the same two parameters of the other five traits in the core collection were all higher than those in the initial collection, implying that the primary core collection has retained a large variation in the initial collection. For the retained ratios, there existed some differences among the six traits. The retained ratio of fruit weight was the highest, with a value of 99.69%; fruit lateral width was in the second place with a value of 97.96%, followed by fruit height (80.00%), ripe stage of fruit (78.57%) and fruit ventral width (70.18%). The value of the soluble solid content was comparably lower; however, the value is still high to 64.35%. These results demonstrated that a large variation exists among the 120 apricot accessions at phenotypic level, and the genetic variation in the core collection can well represent the initial collections, though some accessions were lost in the core collection.

All together, these results confirmed that the sampling strategy we selected was well suitable, and the core collection retained all genetic information of the primary core collection or initial collection, and can represent genetic diversity of the initial collection in apricot.

## DISCUSSION

### The data used to construct core collection

Usually, there are three types of data which were used to establish a crop core collection; that is, passport, characteristic and evaluation data (Brown, 1989). The passport data include some information about the collection site, the eco-physiological condition of origin site, breeding background, principles of taxonomy system, etc; the characteristic data represents the characters of the collection, and includes morphological data, biochemistry data, molecular markers, and so on; the evaluation data includes some agronomic traits about yield, quality, stress resistance, and so on.

When constructing the core collection of a crop, the passport data are usually applied extensively, because those data recorded relative comprehensively, and become the most effective data when combining with other data. As one of the criteria used to study core collection, the morphologic data have some virtues, that is, the method of obtaining data is simple and costs less. However, the morphologic traits usually do not reflect the genetic nature because of some interference caused by the environment or man-made conditions. For example, the plant growth and fruit quality of strawberry were all affected by the temperatures after bloom affect plant growth and fruit quality of strawberry (Wang and Camp, 2000). Alcobendas et al. (2012) reported that exposure to sunlight strongly influenced fruit size, weight and skin color of peach.

Thus, it is important to select the morphologic straits with relative stability to evaluate the germplasm resources. Generally, molecular markers can directly reflect the change of genetic germplasm at DNA level, without interference from the environment, and became the valuable data to evaluate the genetic diversity. It was

**Table 2.** List of 120 accessions and their special characteristics in apricot core collection.

| No. | Germplasm name | Major trait | No. | Germplasm name | Major trait |
|---|---|---|---|---|---|
| 1 | 606Xin | Cold resistance, high acid | 2 | 631Xing | High soluble solids, cold resistance |
| 3 | Anjiana | Flesh with high sugar, cold resistance | 4 | Badouxing | High productivity |
| 5 | Baihuwaina | Early ripening, high soluble solids | 6 | Bairenxing | Ground color of fruit skin is white |
| 7 | Beianhedahuangxing | Large fruit size | 8 | Bianganxing | Early ripening, flesh is firm |
| 9 | Cangjiaxing | Firm and sour flesh | 10 | Caotanmeixing | Flesh is firm |
| 11 | Chuanling | Color on sunny side of one-year-old shoot is red brown | 12 | Chuanzhihongxing | High productivity, flesh is firm, high cold and salt resistance |
| 13 | Dabaixing | Flesh with high sugar | 14 | Dafengxing | Large fruit size, high productivity |
| 15 | Daofuxing | Wild type, high acid | 16 | Dapiantouxing | Fruit in ventral view is clearly asymmetric, fruit scab disease resistance |
| 17 | Dayexing | Attractiveness (1/3 to 1/2 red blush) | 18 | Dongning2hao | Texture of flesh is coarse |
| 19 | Eezhuanzi | Fruit size is very large | 20 | Fakuhebao | The productivity is very high |
| 21 | Fanglingdaxing | Time of fruit maturity is early, fruit size is large | 22 | Guanlaoyelianxing | Light red blush |
| 23 | Guduxing | Wet and hot climate adaptation | 24 | Gushandaxingmei | Early fruit maturity, large fruit size |
| 25 | Guzanxing | Flesh with high soluble solids | 26 | Heiyexing | Flesh with high sugar |
| 27 | Honglianxing | Purple blush | 28 | Honghuomeizi | Deep-red blush, firm flesh |
| 29 | Hongjinzhen | Strong tree vigor, high productivity, very large fruit size | 30 | Hongyuxing | Very large fruit size, sour flesh, broad leaf blade |
| 31 | Huayinxing | Very large fruit size, high productivity | 32 | Huluxing | Processing suitability |
| 33 | Kailidaxing | Very large fruit size, good postharvest characters | 34 | Kailixiaoxing | High productivity |
| 35 | Kangding2hao | High acid | 36 | Kezierkumaiti | High soluble solids, high sugar |
| 37 | Keziximixi | High soluble solids, for drying market | 38 | Kuche3hao | Very high productivity, very small fruit |
| 39 | Kurenhuangkouwai | More dehiscent fruit | 40 | Lajiaoxing | Novel fruit shape (similar to capsicum in shape) |
| 41 | Lintongyinxing | High productivity, large fruit size, high soluble solids | 42 | Luotuohuangxing | Combining very early-ripening with superior fruit quality (size, firmness, sugar) |
| 43 | Maizihuang | Very early ripening | 44 | Meitaoxing | Mediated fruit maturity cultivar |
| 45 | Niujiaohuang | Medium fruit size | 46 | Panxiandashaxing | large fruit size, late fruit maturity |
| 47 | Qingmisha | Very high soluble solids, degree of branching is weak | 48 | Ruanhexing | Degenerate stone |
| 49 | Shaxing1hao | large fruit size, very late fruit maturity | 50 | Shipianhuang | Strong apricot aroma |
| 51 | Shuangrenxing | Very large fruit size, good postharvest characters | 52 | Suanmeizixing | Small fruit, processing cultivar |
| 53 | Tianedan | Very high soluble solids | 54 | Tianrenhuangkouwai | Very high productivity, large fruit size with 1/2 red blush over the skin |
| 55 | Wangjiaxing | Large kernel, pest and disease resistance | 56 | Wanshuxing | Very early fruit maturity, good postharvest characters |
| 57 | Xinjiangshaxing | large fruit size | 58 | Xinshuixing | Good postharvest characters |
| 59 | Xupuxing | Rootstock cultivar | 60 | Yangjiyuanxing | Spreading tree habit, very high productivity |
| 61 | Yinchuantaoxing | Medium fruit size | 62 | Yinghong1hao | Fruit with high pectin |
| 63 | Youyibaixing | Cold resistance | 64 | Youyidaxing | Very high cold resistance, high productivity |
| 65 | Youyiwumingxing | Very high cold and drought resistance, high productivity | 66 | Youyixingmei | High acid, high salt tolerance |
| 67 | Youyiyinbai | Very high cold resistance, high productivity, white petal some with six petals | 68 | Zaohuang | Very high cold resistance, very small fruit size, some blossom with six petals |
| 69 | Zhanggongyuan | Very high productivity, large fruit with large stone, good postharvest characters | 70 | Zhoujiaxing | Flesh with high sugar |
| 71 | Zhupishuixing | Thicker fruit skin, with strong apricot aroma | 72 | Zhuyaoxing | High productivity, firm flesh |
| 73 | Zhuyaozixing | Large fruit size, fruit-setting alternate year | 74 | Chaoxianbaixing | Very large fruit size |
| 75 | Zaoxing | Early fruit maturity, large fruit size, good postharvest characters | 76 | Ningxiataoxing | High fruit quality (large size, succulent flesh with high sugar) |
| 77 | 80A03 | Very large kernel, fruit scab disease resistance | 78 | Baiyubian | Barren soil tolerance |
| 79 | Ganke | For fresh market and kernel consuming | 80 | Huangjianzui | High productivity, medium kernel |

**Table 2.** contd.

| No | Germplasm name | Major traits | No | Germplasm name | Major traits |
|---|---|---|---|---|---|
| 81 | Huzhuazi | Cold resistance, high productivity | 82 | Jiudaomei | Cultivar for flesh market, processing and kernel consuming, fruit-setting alternate year |
| 83 | Kelala | High and regular productivity, kernel is large and sweet | 84 | Longwangmao | Very large kernel |
| 85 | Shanxing | Rootstock cultivar | 86 | Yiwofeng | High productivity |
| 87 | Youyi | High cold resistance, fruit-setting alternate year | 88 | Beilüeshanxing | Green sepal |
| 89 | Chongbanshanxing | Flower with multiple petal | 90 | Chuizhixing | Drooping shoots, narrow leaf blade width |
| 91 | Dayuxingmei | Shoot without thorn, big flower | 92 | Liaomeixing | Flower with multiple petal, very high cold resistance, wild variation of Siberia |
| 93 | Meirenmei | Purple red leaf, interspecific hybridization between cherry and plum | 94 | Yanxingmei | Shoot thorn |
| 95 | Hongfeng | Very early fruit maturity | 96 | Laoshanhongxing | Flesh with high soluble solids, and high sugar, apricot aphid resistance |
| 97 | Longken1hao | Very high cold resistance, high pest and disease resistance | 98 | Longken3hao | High cold resistance, high productivity |
| 99 | Longken5hao | Late ripening, good postharvest characters | 100 | Shiguanzaohong1hao | Very early ripening, self fertility |
| 101 | Xinong25hao | Relative area of over color is large, with red blush, attractiveness, late fruit maturity | 102 | Huiyangbaixing | Fruit scab disease resistance |
| 103 | Jifu | Very high cold resistance | 104 | Jintaiyang | Very early fruit maturity, self fertility |
| 105 | Kaitexing | Very large fruit size, self fertility | 106 | Kaninuo | Small fruit size, sweet-sour flesh with bitter aftertaste |
| 107 | Maonaoxing | Self fertility, flesh with high sugar, good postharvest characters | 108 | Pinghexing | Very small fruit |
| 109 | Xinzhoudashi | Very early fruit maturity, tree habit is spreading | 110 | Zaoju | Specific fruit shape (oblique rhombic) |
| 111 | Yidalixing | Flesh with high sugar | 112 | Meiwuming | High fruit quality (attractiveness, size, firmness, sugar) |
| 113 | Jinfu | High productivity, processing ability | 114 | Changlixingmei | Sour flesh |
| 115 | Hongmeixing | Ground color of skin is red, red flesh | 116 | Jinhuangxingmei | Large fruit size, high productivity, good postharvest characters |
| 117 | Limixing | Late-blooming cultivar | 118 | Longyuanhuangxing | Dwarf tree stature, very high cold resistance |
| 119 | Meiguolixing | Dwarf tree stature, late-blooming | 120 | Qianxianmeixing | High pest and disease resistance, high productivity |

was an effective validating method to investigate the alleles presented in the initial collection whether if also presented in the core collection (Gao et al., 2005). Zhang et al. (2011) constructed the core collection of mulberry using the data of ISSR molecular marker. Whereas, as to the present technology, it may be still labor-intensive and time-consuming to evaluate all the genetic germplasm of one species directly using molecular method (Dong et al., 2003; Zhang et al., 2009a) when the number of crop germplasm is huge.

He et al. (2002) proposed that it is an effective method to construct core collection by morphological data combining molecular marker data. In this study, the core collection established by using different data was evaluated and demonstrated that the core collection based on the single type of data had lower genetic diversity and lower remained ratio. Whereas, the core collection constructed by combining the morphological and agronomical traits data with the molecular marker data had good representative of the initial genetic resources of apricots (Figures 3 and 4). The probably reason was as follows: If only based on the SSR data or other molecular data to construct the core collection, some important germplasms may be lost when the SSR loci selected in the study was not many. Similarly, if only using the morphological data, some germplasms also may be lost because of some interference caused by the environment or man-made conditions.

This study provides some powerful evidence. The best sampling strategy was the cluster sampling method using SSR data combined with MOR data at the rate of 80% among all the sampling methods in this study.

Based on this sampling strategy, in combination with many other germplasm information such as productivity, stress resistance and fruit quality at the same time, 120 accessions were selected as the core collection of apricot (Table 2), which retained 100% alleles in the primary core collection and 100% phenotypical characters, well represented the genetic variance of the primary core collection and the initial collection. Thus, the use of all available information (characteristic and passport data) was found to be valuable for the establishment of apricot core collection. This result was similar to the reports from Diwan in the annual *Medicago* species, which used valuation and passport data to assemble the core collection (Diwan et al., 1995).

**Table 3.** Evaluation of core collection in apricot.

| Parameter | | Fruit weight (g) | Soluble solid content (%) | Fruit development days (d) | Fruit height (cm) | Lateral width (cm) | Ventral width (cm) |
|---|---|---|---|---|---|---|---|
| Original germplasm | Maximum | 133.0 | 29.0 | 190 | 7.2 | 6.7 | 6.6 |
| | Minimum | 2.7 | 6.0 | 50 | 1.7 | 1.0 | 1.7 |
| | Range | 130.3 | 23 | 140 | 5.5 | 5.7 | 4.9 |
| | CV | 44.48 | 20.91 | 15.12 | 15.72 | 15.62 | 16.11 |
| | VPV | 318.18 | 7.50 | 139.01 | 0.40 | 0.35 | 0.42 |
| Primary core collection | Maximum | 133.0 | 23.5 | 160 | 6.3 | 5.5 | 6.6 |
| | Minimum | 3.1 | 7.3 | 50 | 1.9 | 1.5 | 1.8 |
| | Range | 129.9 | 16.2 | 110 | 4.4 | 4.0 | 4.8 |
| | CV | 51.82 | 17.02 | 15.02 | 17.71 | 20.07 | 19.32 |
| | VPV | 457.54 | 4.56 | 138.87 | 0.51 | 0.57 | 0.76 |
| | Retained ratio (%) | 99.69 | 70.44 | 78.57 | 80.00 | 70.18 | 97.96 |
| Core collection | Maximum | 133.0 | 22.5 | 160 | 6.3 | 5.5 | 6.6 |
| | Minimum | 3.1 | 7.7 | 50 | 1.9 | 1.5 | 1.8 |
| | Range | 129.9 | 14.8 | 110 | 4.4 | 4.0 | 4.8 |
| | CV | 53.77 | 17.96 | 15.25 | 18.81 | 20.61 | 18.83 |
| | VPV | 490.74 | 5.12 | 145.06 | 0.57 | 0.61 | 0.57 |
| | Retained ratio (%) | 99.69 | 64.35 | 78.57 | 80.00 | 70.18 | 97.96 |

## The size of core collection and sampling strategy

It is another crucial issue to choose and deal with the suitable data and then decision of the sampling strategy when developing a crop core collection, in detailed, including which sampling methods and which sampling proportion can be used to select the core accessions satisfied the requirements from all the initial accessions. Brown (1989) suggested that about 5 to 10% sample size of the entire collection with an upper limit of 3,000 per species would effectively retain about 70% of the alleles of the entire collection. However, in the study of the core collection of annual *Medicago* species, Diwan proposed that the 5 and 10% sample size core collection were judged insufficient to represent the germplasm collection, probably because the annual *Medicago* species germplasm collection contains many species with very few accessions (Diwan et al., 1995).

The previous study results from many crops suggested that the suitable sampling size of the core accessions was usually about 10 to 30% of the entire collections (Li et al., 2002, 2007; Upadhyaya et al., 2008; Zhang et al., 2009b, 2010). Thus, it is considering that the decision of sampling size should be accorded to the genetic structure and genetic diversity; whereas, could not be uniform simply because the number is different among various crop species, and some special characters occurred with the evolution and the intervening from man-made selection to one crop species. In this study, the size of the core collection (including 120 accessions) was 8% of the initial collection (including 1501 accessions).

For one plant resource, one type of accessions probably many or less, leading to the imbalance in genetic diversity among various accessions. Moreover, there was difference in emphasizing particularly on some genetic structure or genetic diversity from different study departments. As the asymmetric distribution of genetic diversity and the different repetition number of various alleles among the entire accessions, it is necessary to use better sampling strategy to select the core accessions. Based on the sampling strategy, the genetic variance remained as high as possible, and still could not change the genetic structure of the initial collection. According to present literatures, it was optimal to use stratified cluster

sampling methods (Hu et al., 2000; Jansen and van Hintum, 2007; Zhang et al., 2009b). Li et al. (2002) conducted the study on sampling schemes for the establishment of core collection of rice landraces in Yunnan and suggested that clustering sampling methods was invariably better than non-clustering, whenever at the same grouping principles or at the same sampling proportion within group. Similar to this result, clustering sampling methods is much better than random sampling in this study, and cluster sampling of SSR combined with MOR at the rate of 80% was the best sampling strategy used to construct apricot core collection. There were few reports about the assemblage of core collection in horticultural crop species especially in fruit plants germplasm, and thus, there were less available methods. In this study, the molecular marker data was first used to establish apricot core collection, 120 accessions was selected from 1501 initial collections, which enable the utilization to be much convenient. However, as apricot species is a woody plant, it is necessary to validate the practicality of the core collection by growing judge.

## ACKNOWLEDGEMENTS

This research was financially supported by Beijing Nova Program (No. 2009B27), Municipal Natural Science Foundation of Beijing (No. 6122011) and Special Fund for Agro-scientific Research in the Public Interest (No. 201003058).

### REFERENCES

Alcobendasa R, Mirás-Avalosb JM, larcóna JJ, Pedreroa F, Nicolása E (2012). Combined effects of irrigation, crop load and fruit position on size, color and firmness of fruits in an extra-early cultivar of peach. Sci. Hortic. 142:128-135.

Brown AHD (1989). Core collection:a practical approach to genetic resources management. Genome 31:8l8-824.

Cao XY, Feng JR, Wang DJ, Sun JL, Lu XY, Liu HF (2012). Primary style protein expression in the self-incompatible/compatible apricot by the 2D-DIGE technique. Gene 503:110-117.

Dirlewanger E, Cosson P, Tavaud M, Aranzana MJ, Poizat C, Zanetto A, Arús P (2002). Development of microsatellite markers in peach [*Prunus persica* (L.) Batsch] and their use in genetic diversity analysis in peach and sweet cherry (*Prunus avium* L.). Theor. Appl. Genet. 105:127-138.

Diwan N, McIntosh MS, Bauchan GR (1995). Methods of developing a core collection of annual Medicago species. Theor. Appl. Genet. 90:742-745.

Dong YS, Cao YS, Zhang XY, Liu SC, Wang LF, You GX, Pang BS, Li LH, Jia JZ (2003). Establishment of candidate core collections in Chinese common wheat germplasm. J. Plant Genet. Resour. 4(1):1-8 (In Chinese).

Doyle JJ, Doyle JL (1987). A rapid DNA isolation procedure from small quantities of fresh leaf tissues. Phytochem. Bull.19:11-15.

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

Gao ZH, Zhang Z, Han ZH, Fang JG (2005). Development and evaluation of core collection of Japanese apricot germplasms in China. Sci. Agric. Sin. 38:363-368 (In Chinese).

Grenier C, Branel-Cox PJ, Noirot M, Rao KEP, Hamon P (2000). Assessment of genetic diversity in three subsets constituted from the ICRISAT sorghum collection using random vs non-random sampling procedures A. Using morpho-agronomical and passport data. Theor. Appl. Genet. 101:190-196.

Hagen LS, Chaib J, Fady B, Decroocq V, Bouchet JP, Lambert P, Audergon JM (2004). Genomic and cDNA microsatellites from apricot (*Prunus armeniaca* L.). Mol. Ecol. Notes 4:742-745.

He YT, Tu JX, Fu TD, Li DR, Chen BY (2002). Preliminary development of core collection of *Brassica campestris* in Shanxi province. Chin. J. Oil Crop Sci. 24(1):6-9 (In Chinese).

Hu J, Zhu J, 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.

Jansen J, van Hintum TH (2007). Genetic distance sampling:a novel sampling method for obtaining core collections using genetic distances with an application to cultivated lettuce. Theor. Appl. Genet. 114:421-428.

Li YX, Li TH, Zhang HL, Qi YW (2007). Sampling strategy for a primary core collection of peach (*Prunus persica* (L) Batsch.) germplasm. Eur. J. Hortic. Sci. 72 (6):268-274.

Li ZC, Zhang HL, Zeng YW, Yang ZY, Shen SQ, Sun CQ, Wang XK (2002). Studies on sampling schemes for the establishment of core collection of rice landraces in Yunnan, China. Genet. Resour. Crop Evol. 49:67-74.

Liu N, Liu WS (2006). Descriptors and data standard for apricot (*P. armeniaca* Mill.) [M]. Beijing: China Agricultural Press. (in Chinese).

Lopes MS, Sefc KM, Laimer M, Câmara Machado ADA (2002). Identification of microsatellite loci in apricot. Mol. Ecol. Notes 2:24-26.

Sosinski B, Gannavarapu M, Hager LD, Beck LE, King GJ, Ryder CD, Rajapakse S, Baird WV, Ballard RE, Abbott AG (2000). Characterization of microsatellite markers in peach [*Prunus persica* (L.) Batsch]. Theor. Appl. Genet. 101:421-428.

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

Testolin R, Marrazzo, T, Cipriani G, Quarta R, Verde I, Dettori MT, Pancaldi M, Sansavini S (2000). Microsatellite DNA in peach (*Prunus persica* L. Batsch) and its use in fingerprinting and testing the genetic origin of cultivars. Genome. 43:512-520.

Upadhyaya HD, Pundir RPS, Gowda CLL, Gopal Reddy V, Singh S (2008). Establishing a core collection of foxtail millet to enhance the utilization of germplasm of an underutilized crop. Plant Genet. Resour. 7(2):177-184.

Wang YS, Camp MJ (2000). Temperatures after bloom affect plant growth and fruit quality of strawberry. Sci. Hortic. 85:183-199.

Wang YZ (1998). Recommendation of apricot cultivars for commercial growing in China. Hort. Sci. (Praha). 25:121-124.

Wang YZ, Zhang JH, Ning N, Sun HY, Yang L (2011). Construction and evaluation of a primary core collection of apricot germplasm in China. Sci. Hortic. 128:311-319.

Yao QL, Fan P, Zou SX (2008). Constructing a core collection for maize (*Zea mays* L.) landrace from Wuling Mountain region in China. Agric. Sci China. 7(12):1423-1432.

Zhang CY, Chen XS, Zhang YM, Yuan ZH, Liu ZC, Wang YL, Lin Q (2009b). Method of constructing core collection for *Malus sieversii* in Xinjiang, China using molecular markers. Agric. Sci China. 8:276-284.

Zhang J, Wang Y, Zhang XZ, Li TH, Wang K, Xu XF, Han ZH (2010). Sampling strategy to develop a primary core collection of apple cultivars based on fruit traits. Afr. J. Biotechnol. 9:123-127.

Zhang L, Zhao WG, Chen JB, Huang Y, Shen XJ, Liu L, Qiang S (2011). Analysis of genetic diversity and construction of core collection of local mulberry varieties from Shanxi Province based on ISSR marker. Afr. J. Biotechnol. 10:7756-7765.

Zhang YF, Zhang QL, Yang Y, Luo ZR (2009a). Development of Japanese persimmon core collection by genetic distance sampling based on SSR markers. Biotechnol. Biotec. Eq. 23(4):1474-1478.