Europ.J.Hort.Sci., **72** (6). S. 268–274, 2007, ISSN 1611-4426. © Verlag Eugen Ulmer KG, Stuttgart

Sampling Strategy for a Primary Core Collection of Peach (*Prunus persica* (L.) Batsch.) Germplasm

Y. X. Li[1)], T. H. Li[1)], H. L. Zhang[2)] and Y. W Qi[2)]
([1)]Department of Fruit Science, College of Agronomy, Biotechnology, China Agricultural University, Beijing, and [2)]College of Crop Science, China Agricultural University, Beijing, P. R. China)

Summary

Sampling strategies were evaluated to select for a primary core collection of peach (*Prunus persica* (L.) Batsch.) that would be the representative of the base collection of peach cultivars at the Beijing's National Peach Germplasm Resources Repository. The optimal sampling strategy was established based on the coverage of morphological and agronomic characters in 18 taxonomic groups. In order to determine the most representative core collection, we used a simulation approach to identify the optimal sampling scheme. Nine entire sampling ratios from 10 % to 90 % were tested. The sampling scheme was conducted at two levels, first to divide the collection into biologically relevant groups and then to determine the appropriate core selections within groups. The results indicated that 10 % was the appropriate entire sampling ratio from the base collection. The optimal sampling scheme was to divide all the accessions into different variety groups, and then determine the sampling proportion within groups based on its genetic diversity (G) using a random sampling method. Using this sampling strategy, a primary core collection with 56 accessions was selected from a base collection of 558 accessions of different peach cultivars. The primary core collection represented 99.49 % of the ratio of phenotype retained (RPR) of the base collection, and its index of genetic diversity (I) was 0.63, which was significantly higher than 0.59 of the base collection. The evaluation results indicated that the primary core collection was able to represent the initial germplasm collection in genetic diversity and heterogeneity.

Key words. Core collection – sampling scheme – grouping principles – sampling proportion within groups – rank analyses

Introduction

China is the world's largest conservator of peach genetic diversity of the world. About 1000 accessions of 800 cultivars of peach germplasm have been collected and conserved in China's three National Peach Germplasm Resources Repositories (Beijing, Nanjing and Zhengzhou) (Wang and Zhuang 2001). The peach genetic resource collection in these repositories will increase rapidly with the enhancement of international exchange and the acceleration of breeding in China. Because peach trees are big and perennial, germplasm preservation in the field can be very costly. Insects, diseases and unfavourable environmental conditions also make the peach germplasm collection vulnerable.

A core collection is a minimal subset of the entire collection and has most of the genetic diversity of a species (Frankel 1984; Brown 1989). The construction of core collections can effectively overcome and alleviate the difficulty of maintaining exhaustive collections by reducing genetically redundant materials. Core collections will make it easier for germplasm conservation, evaluation, innovation and utilization. A good core collection should be able to provide for a wide spectrum of genetic bases and strong compatibilities (Huaman et al. 2000). Core collections of many plant species have been established around the world, including wheat (Spagonletti and Quaslset 1991; Dong et al. 2003), alfalfa (Diwan et al. 1995), soybean (Perry et al. 1991; Wang et al. 2004), peanut (Holbrook and Anderson 1995; Upadhyaya et al. 2002, 2003), rice (Li et al. 2002), apple (Hokanson et al. 1998) and Japanese apricot (Gao et al. 2005). However, studies to produce a core collection of peach had not yet been reported.

The origin, geographical distribution, ecotypes and breeding methods all affect the genetic diversity of peach and the genetic structure of the biodiversity in the current germplasm collection. Li et al. (2003) has proposed a two-step strategy for establishment of a core collection after researching the rice germplasm resources in China. Step one is selection of a primary core collection which includes optimisation of a sampling strategy based on agronomic characters to determine the optimal entire sampling ratio and sampling scheme, the latter includes the grouping principles and the sampling proportions and sampling methods within groups. Step two is to build the core collection based on further classification of the collections according to phenotypical characteristics, isoenzyme markers, and DNA molecular data. The principles

for establishing the core collection are to reduce the variations at high frequencies, to increase the proportion of rare variations in the core collection and ensure that the main biological types within each group will be included (SUN et al. 2001; LI et al. 2003). As a results, the diversity within groups is minimized, while diversity between groups is maximized, and this will guarantee that the core collection represents the genetic diversity of the base collection.

Our objective was to select an optimal sampling strategy. The strategy included an optimal entire sampling ratio from the base collection and an appropriate sampling scheme. This was used to identify the reliability of the primary core collection and to establish an ideal primary core collection of peach that would represent the diversity of the base collection.

Materials and Methods

558 accessions of peach cultivars conserved in Beijing's National Peach Germplasm Resources Repository were used in this study. The basic data was obtained from the Catalogue of Germplasm Resources of Fruit Trees (ANONYMOUS 1993, 1998). The characters included the presence or absence of pollen, regions of origin and fruit characteristics, such as fruit form, size, average weight per fruit, fruit fuzziness, ground colour, colour of fruit, fruit texture, colour of pulp, fruit flavour and juice, pit glutinosity, soluble solids content, soluble sugar, soluble acid, vitamin C and fruit ripening stage. Most of the characters studied were adjusted or standardized according to Genetic Evaluation of Main Characters in Peach (WANG and ZHUANG 2001). Quantitative characters were quantified into different categories, where the distance between two neighbouring categories was 0.5 standard deviation.

Sampling strategy for the primary core collection

The sampling strategy for the establishment of the primary core collection of peach included the determination of an entire sampling ratio from the base collection and a reliable sampling scheme. The entire sampling ratio would determine the size of the primary core collection. Nine sampling ratios, 10, 20, 30, 40, 50, 60, 70, 80 and 90 %, were tested. In order to include all the main biological types in the primary core collection, at least one accession from each group was sampled.

The sampling scheme was conducted at two levels, (1) grouping principles, including a) the region of origin, accessions were separated by whether the cultivar was bred or originated in China, or introduced from foreign countries; b) cultivated region, accessions were separated by the region of the Northwest China's Arid Area, the Northeast China's Cold Zone, the North China Plain, the Yangze River Valley, the Yunan-Guizhou Plateau Area, the Qinghai-Tibet Plateau Area, and Tropical Zone in Southern China (CHENG and HUANG 2004); c) ecotype, accessions were divided into the Northern China population, the Southern China population, and European population (WANG and ZHUANG 2001); d) variety group, accessions were divided into crisp peach, sweet peach, honey peach, flat peach, nectarine, and yellow fleshed peach variety groups according to the WANG and LU (1990) method. e) Completely random sampling (non-grouping) for all the accessions. (2) sampling proportion within each group, four different approaches were used, including a) proportion of genetic diversity (G), where the number of accessions to sample from each group was proportional to the group's genetic diversity (G = genetic diversity index of each group / sum of genetic diversity indexes of every group in the whole population under a certain grouping principle); b) simple proportion (P), referred to a certain fixed percent of the accessions from each group; c) proportion of square root (S), where the weight of each group in the core was proportional to the square root of group's size (S= square root of each group / sum of square roots of every group in the whole population under a certain grouping principle); and d) proportion of logarithm (L), where the number of accessions sampled from each group was proportional to the logarithm of group's size (L= logarithm of each group / the sum of the logarithms of every group in the whole population under a certain grouping principle). After the number of accessions to be sampled from each group was defined, the random sampling method was used within groups. There were a total of 17 schemes based on the combination of the above two levels. Each sampling scheme was performed with 5 replicates to reduce the random errors.

Six parameters were used to test the genetic diversities of the primary core collections constructed by different sampling schemes (ORTIZ et al. 1998; UPADHYAYA et al. 2002; ZHANG et al. 2003): index of genetic diversity (I) is a measure of both richness and evenness of variation of a population in genetic studies; ratio of phenotype retained (RPR), is the percentage of the number of core collection phenotypes out of the total number of phenotypes in the base collection, that assesses the richness of variation in the core collection; coefficient of variation (CV) and variance of phenotypic value (VPV), the parameters to test the distribution of the phenotypes, is a measure of heterogeneity and can be used to evaluate the evenness of the population; deviation of phenotypic minimum (D_{min}) and deviation of phenotypic maximum (D_{max}) are the deviation of the core collection from the base germplasm population and can represent the range of genetic variation. The parameters were computed as follows:

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

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

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

$$D_{min} = \frac{STD_i(Max_i - Max_{i0})}{N}$$

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

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

where p_{ij} shows the frequency of phenotype j for character i; N shows the number of characters in use; M_i shows the number of phenotype of character i in core collection; M_{i0} shows the number of phenotype of character i in base collection; Min_i shows minimum phenotype value of character i in core collection; Max_{i0} shows maximum phenotype value of character i in base collection; Min_{i0} shows minimum phenotype value of character i in base collection; Max_i shows maximum phenotype value of character i in core collection; X_{ij}, shows the phenotype of character i for accession j; X I shows the mean of phenotype for character i; and STD_i shows standardization of characters i. Standardization refers to normalization of all the data into the standard normal distribution. The normalized value = [(original value − mean of all the data before normalization)/standard deviation (SD)] of all the data before normalization.

The quantitative data were used in both a quantitative and a qualitative scale. The former scale was used to estimate parameters D_{min}, D_{max}, VPV and CV; and the latter to estimate parameters I and RPR by converting the quantitative data into a qualitative scale.

Each primary core collection was evaluated for their ability to represent the base collection according to these parameters using a Foxpro 6.0 system. Statistical analysis of the variances for the 6 parameters was performed with SPSS 11.0 software. Variance analysis was employed to compare the differences of the 6 parameters among different sampling ratios from the base collection, different grouping principles and different sampling proportions within groups. The sampling schemes were ranked based on the comprehensive evaluation of the 6 parameters. Ranking was based on the Variance values and scored at a statistical difference level of 0.05. The same ranking score implied that there was no difference, whereas different scores, arranged from low to high, implied better to inferior merit.

Evaluation of the primary core collection

By comparing the RPR and I of the investigated characters of the base collection and primary core collection, and the analysis of D_{max}, D_{min}, Range, CV, VPV, and proportion of range (PR) of 6 quantitative characters (duration of fruit development, average fruit weight, content of soluble solids, soluble sugar content, soluble acid content, vitamin C), the representative of the primary core collection was evaluated. The PR of quantitative characters was computed as follows:

$$PR_i = \frac{Max_i - Min_i}{Max_{i0} - Min_{i0}} \times 100\ \%$$

Where PR_i was the proportion of range for character i, Max_i, Min_i, Max_{i0}, and Min_{i0} were the same as above.

Results

Studies of the optimal entire sampling ratio

A total of 153 candidate primary core collections were conceived using 9 entire sampling ratios from the base collection and 17 sampling schemes. Using 6 parameters to compare the different entire sampling ratios, only I and VPV reached the significant level ($P<0.05$) (data not shown). When the entire sampling ratio was 10 %, the primary core collections had the highest average values of I (0.6027) and VPV (0.2111); the ranks of both I and VPV, and their average were the lowest at 1.0 (Table 1). When the phenotypes were absent, the P_{ij} =0, these data were excluded from further calculation. The results suggested that the primary core collections established under the entire sampling ratio of 10 % have the highest richness and evenness of genetic variation, and the minimum degree of genetic redundancy. Therefore, 10 % would be the best entire sampling ratio for the primary core collection of peach.

Studies of grouping principles and sampling proportions within groups

When the entire sampling ratio from the base collection was fixed at 10 % and the random sampling method was used, different grouping principles and different sampling proportions within each group were compared for their coverage of genetic diversity among the primary core collections. The results showed that grouping the base collection by variety group had the best effect (average rank 1.00, Table 2), followed by grouping according to region of origin and ecotype (average rank 1.50 and 1.67, respectively). Grouping by cultivated region had the poorest effect (average rank 2.50), indicating that this was not a proper criterion. D_{min} was the only parameter that showed a significant difference among the 4 different sampling proportions within groups (G, P, S, and L). As D_{min} indicates the dispersion between the minimum value of each character in the core collection and in the base collection, it carries very little genetic information. Therefore, the effectiveness of these 4 sampling proportions within groups could not be determined directly. So further study was needed to compare the interactions of grouping principles and sampling proportions within groups.

Studies of the sampling schemes for the primary core collection

Table 3 showed the interactions between the different grouping principles and the 4 sampling proportions within groups (G, P, S, and L) under the conditions of the entire sampling ratio of 10 % and random sampling from each group. 17 primary core collection schemes were established by the interactions between 4 grouping principles and 4 proportions within groups, and one of non-grouping with simple proportion. Two sampling schemes, i.e. i) grouped the accessions into different variety groups and then weighted the groups based on the proportions of genetic diversity (variety group + G), and ii) grouped the accessions into different variety groups and then weighting the groups based on the proportions of square root (variety group + S), had the lowest average rank of 1.33 and should be the best sampling schemes. Further comparison of these two sampling schemes, (variety group + G) was significantly advantageous to (variety group + S) considering the RPR and VPV values of the primary core collections. The results indicated that the (variety group + G) scheme produced a primary core collection that preserved more genetic variations, little genetic redundancy and even distribu-

Table 1. Comparison of average values of index of genetic diversity (I), variance of phenotypic value (VPV) and ratio of phenotype retained (RPR); and rank analyses of I and VPV under different entire sampling ratios.

Entire sampling ratio (%)	I	VPV	Average rank	RPR (%)
10	0.6027 (1.0)	0.2111 (1.0)	1.00	97.40
20	0.5981 (2.0)	−0.0637 (2.5)	2.25	99.07
30	0.6000 (2.0)	0.0489 (2.0)	2.00	99.42
40	0.5996 (2.0)	−0.0449 (2.5)	2.25	99.74
50	0.5992 (2.0)	−0.1198 (3.0)	2.50	99.94
60	0.5994 (2.0)	−0.0733 (2.5)	2.25	100.00
70	0.5983 (2.0)	−0.1692 (3.5)	2.75	100.00
80	0.5985 (2.0)	−0.2138 (3.5)	2.75	100.00
90	0.5984 (2.0)	−0.2523 (4.0)	3.00	100.00

Note: The values in the brackets were the ranks of each testing parameter. I=Index of genetic diversity, VPV=Variance of phenotypic value, and RPR=Ratio of phenotype retained.

Table 2. Values and rank analyses of 6 testing parameters of primary core collections established by different grouping principles or different sampling proportions within groups.

		Testing parameter						Average rank
		I	RPR	D_{min}	D_{max}	VPV	CV	
Grouping principle	Variety group	0.6101 (1.0)	0.9910 (1.0)	1.0937 (1.0)	1.2808 (1.0)	0.7345 (1.0)	0.4994 (1.0)	1.00
	Region of origin	0.6033 (1.0)	0.9769 (2.0)	1.1489 (1.0)	1.8747 (2.0)	0.2254 (1.5)	0.1403 (1.5)	1.50
	Ecotype	0.6065 (1.0)	0.9731 (2.0)	1.5693 (1.0)	1.8105 (2.0)	0.2504 (1.5)	0.0843 (2.5)	1.67
	Cultivated region	0.5911 (2.0)	0.9551 (3.0)	3.3124 (2.0)	2.7004 (3.0)	0.3660 (2.0)	0.4845 (3.0)	2.50
Sampling proportion within groups	G	0.6008 (1.0)	0.9692 (1.0)	1.8573 (2.0)	1.6651 (1.0)	0.2979 (1.0)	0.0058 (1.0)	1.17
	P	0.6039 (1.0)	0.9718 (1.0)	1.3721 (1.0)	2.009 (1.0)	0.2108 (1.0)	0.0796 (1.0)	1.00
	S	0.6045 (1.0)	0.9756 (1.0)	2.0723 (2.0)	2.1031 (1.0)	0.1783 (1.0)	0.1440 (1.0)	1.17
	L	0.6018 (1.0)	0.9795 (1.0)	1.8226 (1.5)	1.8896 (1.0)	0.1573 (1.0)	0.1469 (1.0)	1.08

Note: The values in the brackets were the ranks of each testing parameter. I=Index of genetic diversity, RPR=Ratio of phenotype retained, D_{min}= Deviation of phenotypic minimum, D_{max}=Deviation of phenotypic maximum, VPV=Variance of phenotypic value, CV=Coefficient of variation, G= Proportion of genetic diversity, P=Simple proportion, S= Proportion of square root, L= Proportion of logarithm.

tion of each character. Although (variety group + S) scheme caused lower D_{max} value, the D_{max} represents the deviation between the maximum value of each character in the core collection and in the base collection, which only indicated the variation spectrum. Considering the core collection should keep the most of the genetic variability of the whole gene bank, the best sampling scheme to establish a peach primary core collection was to divide all the accessions into different variety groups, and then determine the sampling proportion from each variety group according to proportion of genetic diversity (G) using random sampling method. Therefore, using this optimal sampling scheme and 10 % of the entire sampling ratio, a primary core collection of peach with 56 accessions was selected from a base collection of 558 accessions.

Evaluation of the primary core collection

The primary core collection (Table 4) was evaluated for its ability to represent the base collection in 18 morphological and agronomic characters. The RPR of the primary core collection was 99.49 %, retaining basically all the variation of the base collection. The value of I of the

Table 3. Values and rank analyses of 6 testing parameters under 17 different sampling schemes for primary core collection.

Sampling scheme		Testing parameter						Average
Grouping principle	Sampling proportion within groups	I	RPR	D_{min}	D_{max}	VPV	CV	rank
Variety group	G	0.6137 (1.0)	0.9949 (1.0)	0.6903 (2.0)	1.3396 (2.0)	1.1164 (1.0)	0.9073 (1.0)	1.33
	P	0.6094 (1.5)	0.9949 (1.0)	1.3079 (2.0)	1.2202 (1.0)	0.7176 (1.5)	0.4093 (1.5)	1.42
	S	0.6124 (1.0)	0.9795 (1.5)	0.9692 (2.0)	1.2682 (1.0)	0.9335 (1.5)	0.9131 (1.0)	1.33
	L	0.6050 (2.0)	0.9949 (1.0)	0.6903 (2.0)	1.2953 (1.5)	0.1705 (2.0)	0.2321 (2.0)	1.75
Region of origin	G	0.5972 (2.5)	0.9590 (3.0)	1.5890 (2.5)	1.3401 (2.0)	0.5963 (1.5)	0.1970 (2.0)	2.25
	P	0.6045 (2.0)	0.9744 (2.0)	0.2970 (1.0)	1.4277 (2.5)	0.5446 (1.5)	0.4912 (1.5)	1.75
	S	0.5986 (2.5)	0.9949 (1.0)	1.7303 (2.5)	2.6936 (5.5)	0.3315 (2.0)	−0.1970 (2.0)	2.58
	L	0.6131 (1.0)	0.9795 (1.5)	1.5733 (2.5)	2.0373 (3.0)	0.0924 (2.0)	0.0645 (2.0)	2.00
Ecotype	G	0.6043 (2.0)	0.9744 (2.0)	1.6966 (2.5)	1.5054 (2.5)	0.4188 (2.0)	0.0347 (2.0)	2.17
	P	0.6097 (1.5)	0.9590 (3.0)	1.4162 (2.0)	2.1217 (3.0)	0.0806 (2.0)	0.0256 (2.0)	2.25
	S	0.6073 (1.5)	0.9846 (1.5)	1.4583 (2.0)	1.9073 (3.0)	0.1526 (2.0)	0.3742 (2.0)	2.00
	L	0.6046 (2.0)	0.9744 (2.0)	1.7060 (2.5)	1.7075 (3.0)	0.6546 (1.5)	0.0973 (2.0)	2.17
Cultivated region	G	0.5878 (3.5)	0.9487 (3.5)	3.4532 (4.5)	2.4752 (4.0)	0.9399 (3.0)	1.0929 (3.0)	3.58
	P	0.5920 (3.0)	0.9590 (3.0)	3.0614 (4.0)	3.2647 (6.0)	0.4998 (2.5)	0.5565 (2.5)	3.50
	S	0.5998 (2.5)	0.9436 (4.0)	4.1313 (5.0)	2.5435 (5.0)	0.2640 (2.0)	0.2288 (1.5)	3.33
	L	0.5847 (4.0)	0.9692 (2.5)	2.6036 (3.5)	2.5181 (4.5)	0.2884 (2.0)	0.5174 (2.5)	3.17
Non-grouping	P	0.5937 (3.0)	0.9487 (3.5)	1.3046 (2.0)	1.1730 (1.0)	0.0980 (2.0)	0.0105 (2.0)	2.25

Note: The values in the brackets were the ranks of each testing parameter. I=Index of genetic diversity, RPR=Ratio of phenotype retained, D_{min}= Deviation of phenotypic minimum, D_{max}=Deviation of phenotypic maximum, VPV=Variance of phenotypic value, CV=Coefficient of variation, G= Proportion of genetic diversity, P=Simple proportion, S= Proportion of square root, L= Proportion of logarithm.

core was 0.63, which was significantly higher than 0.59 of the base collection ($P<0.05$). This implied that the richness and evenness of the primary core collection estimated by parameter I improved as compared with that of the base collection. Comparison of CV, VPV, and PR of 6 quantitative characters (duration of fruit development, average fruit weight, soluble solids content, soluble sugar content, soluble acid content and vitamin C) was conducted between the primary core collection and the base collection. The results showed that although CV of vitamin C was lower, CV of the other 5 characters and VPV of all these 6 characters in the primary core collection were higher than those of the base collection, indicating that the primary core collection had good heterogeneity. The PR of average weight per fruit was the highest at c.a. 91.85 %, followed by that of the duration of fruit development (87.41 %), soluble solids (82.96 %), and soluble acid (82.35 %). The PRs of soluble sugar and vitamin C were relatively low, but still around 70 %. Therefore, the primary core collection was able to represent the base germplasm collection in genetic diversity and heterogeneity.

Table 4. Evaluation of peach primary core collection.

Character	Base germplasm					Primary core collection					PR
	Max	Min	Range	CV	VPV	Max	Min	Range	CV	VPV	(%)
Duration of fruit development (d)	192.0	57.0	135.0	23.18	649.43	175.0	57.0	118.0	26.11	890.72	87.41
Average weight per fruit (g)	353.6	12.5	341.1	35.48	2748.09	325.8	12.5	313.3	40.38	2892.18	91.85
Soluble solids content (%)	20.0	6.5	13.5	17.91	4.54	20.0	8.8	11.2	20.21	6.54	82.96
Soluble sugar content (%)	18.4	4.2	14.2	20.00	2.97	14.0	4.2	9.8	20.51	3.53	71.68
Soluble acid content (%)	1.8	0.1	1.7	51.22	0.08	1.5	0.1	1.4	55.20	0.09	82.35
Vitamin C (mg 100 g^{-1})	28.5	2.7	25.8	39.37	13.58	20.6	2.7	17.9	36.94	15.25	69.38

Note: Range=Max-Min, Max=Value of maximum, Min=Value of minimum, CV=Coefficient of variation, VPV=Variance of phenotypic value, PR= Proportion of range.

Discussion

Determination of the optimal entire sampling ratio

The objective of constructing a core collection is to maximally represent the diversity of a germplasm collection with minimum redundancy. In the previous studies of constructing core collections on different plant species, the entire sampling ratios were between 5 % and 30 % of the entire germplasm collection, or the total number of accessions was below 3 000 (Brown 1989). The size of the population and the genetic structure of the base collection are the key factors that affect the entire sampling ratio to establish a core collection (Li et al. 2003; Sun et al. 2001). The ratio of a core collection to its base collection can be smaller when the size of the base collection is larger or the base collection has a smaller genetic diversity. Otherwise, the entire sampling ratio should be increased. Brown (1989) expected that a sample size of about 10 % would retain 70 % of the alleles present in the whole collection with 95 % confidence. This study was conducted on a relatively small number of germplasm. The primary core collection is the preliminary compression of the base germplasm collection. The minimum entire sampling ratio of 10 % from the whole population was selected, in order to maximize the inclusion of genetic diversity and to avoid any misleading effect by the phenotypic characters. The results showed that the RPR retained in the primary core collection reached over 97 % when the entire sampling ratios were from 10 % to 90 %. When the ratio was 10 %, the primary core collection had RPR of 97.40 %, the highest I (0.6027) and VPV (0.2111) values and the lowest average rank (1.0). This indicated that the primary core collection of peach obtained under the entire sampling ratio of 10 % had high richness and evenness of genetic variation with the minimum degree of genetic redundancy. According to the concept of a core collection, we proposed that 10 % should be the optimal entire sampling ratio for the primary core collection of peach.

Determination of sampling scheme

The sampling scheme included the determination of grouping principles and sampling proportions within groups. The optimisation of the sampling scheme was conducted using 10 % as the entire sampling ratio from the base collection. In general, the sampling schemes that grouped the accessions first were better than the one without grouping (Table 3). The optimal grouping principle was confirmed by rank analyses of the 6 testing parameters. The effectiveness of 4 sampling proportions within groups could not be distinguished (Table 2).

There was an obvious interactive effect between the grouping principle and sampling proportion within groups (Table 3). When grouping accessions based on variety group, sampling proportions according to G and S produced better sampling effects. However, if grouping by region of origin, using the simple proportion (P) to sample from each group was better than using the other 3 sampling proportions within groups. Therefore, an adequate classification of accessions into related groups should be performed prior to the selection of an optimal sampling proportion within groups (Malosetti and Abadie 2001).

Peach germplasm classification has not yet been standardized. The existing methods include classifying the peach cultivars based on geographic origin + ecotype, fruit characters, or even based on the cultivation characteristics like tree vigour or the presence or absence of pollen (Cheng and Huang 2004). Wang and Lu (1990) put forward the stratified standard of peach accessions when they were working on the evolution and classification of peach germplasm based on isozyme data. The peach accessions were divided into six variety groups, i.e. crisp peach, sweet peach, honey peach, flat peach, nectarine, and yellow fleshed peach. This classification method considered the fruit morphology, fruit texture and also involved the regions of origin that affect many important biological characteristics. This covered many sets of important biological characteristics and high polymorphism. Using this method to classify the peach accessions

can effectively stratify the entire original germplasm into non-overlapping subsets, and adequately reflect the genetic structure of peach germplasm. It was proved in our study that grouping by variety group was the best principle in the sampling scheme (Table 3).

Sampling proportions within each group should be determined in accordance with the genetic diversity within each variety group (CHANDRA et al. 2002; LI et al. 2003). The information of the phenotypic diversity of the base germplasm of peach used in this study is relatively integrated. Due to this, grouping by variety group and then weighting the groups based on the proportion of genetic diversity became the best optimal sampling scheme. Compared with the base germplasm collection, the primary core collection based on this sampling scheme using a 10 % entire sampling ratio was able to adequately represent the genetic spectrum and maximally capture its genetic diversity.

In conclusion, 10 % was the appropriate ratio of the primary core collection to the base collection. The optimal sampling scheme was to divide all the accessions into different variety groups, and then determine the sampling proportion within each group based on its genetic diversity (G) using the random sampling method. Using this sampling strategy the primary core collection of 56 accessions from the base collection of 558 accessions of peach cultivars was acquired. The evaluation results showed that this primary core collection was able to adequately represent the base germplasm collection in genetic diversity and heterogeneity.

Acknowledgments

This research was funded by the China Postdoctoral Science Foundation and the Program for New Century Excellent Talents in University of China. We appreciate the work of Prof. Q. Jiang and Dr. J.B. Zhao (Beijing's National Peach Germplasm Resources Repository, China) for expending a lot of effort to collect and take care of the plant materials. We also thank Prof. Z.C. Li (China Agricultural University, China) for his advice on statistical analysis; and Mr. Ethan Johnson (The Holden Arboretum, Kirtland, Ohio, U.S.A.) and Ms. J. Wang (Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, China) for their critical review of the manuscript.

References

ANONYMOUS 1993: The Catalogue of Germplasm Resource of Fruit Tree (Vol. 1). Institute of Fruit Science, Chinese Academy of Agricultural Sciences. Chinese Agricultural Press, Beijing, 56–89.

ANONYMOUS 1998: The Catalogue of Germplasm Resource of Fruit Tree (Vol. 2). Institute of Fruit Science, Chinese Academy of Agricultural Sciences. Chinese Agricultural Press, Beijing, 24–31.

BROWN, A.H.D. 1989: Core collection: a practical approach to genetic resources management. Genome **31**, 818–824.

CHANDRA, S., Z. HUAMAN, S. HARI KRISHNA and R. ORTIZ 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.

CHENG, Z.P. and H.W. HUANG 2004: Genetic diversity and genetic structure of different groups in *Prunus persica* revealed by RAPD analysis. J. Wuhan Bot. Res. **22**, 27–32.

DIWAN, N., M.S. MCINTOSH and G.R. BAUCHAN 1995: Methods of developing a core collection of annual *Medicago* species. Theor. Appl. Genet. **91**, 755–761.

DONG, Y.S., Y.S. CAO, X.Y. ZHANG, S.C. LIU, L.F. WANG, G.X. YOU, B.S. PANG, L.H. LI and J.Z. JIA 2003: Establishment of candidate core collections in Chinese common wheat germplasm. J. Plant Genet. Res. **4**, 1–8.

FRANKEL, O.H. 1984: Genetic perspectives of germplasm conservation. In: ARBER, W., K. ILIMENSEE, W.J. PEACOCK and P. STARLINGER (eds.): Symposium on genetic manipulation: Impact on man and society. Cambridge University Press, Cambridge, 161–170.

GAO, Z.H., Z. ZHANG, Z.H. HAN and J.G. FANG 2005: Development and evaluation of core collection of Japanese apricot germplasms in China. Sci. Agr. Sinica. **38**, 363–368.

HOKANSON, S.C., A.K. SZEWC-MCFADDEN, W.F. LAMBOY and J.R. MCFERSON 1998: Microsatellite (SSR) markers reveal genetic identities, genetic diversity and relationships in *Malus×domestica* Borkh. core subset collection. Theor. Appl. Genet. **97**, 671–683.

HOLBROOK, C.C. and W.F. ANDERSON 1995: Evaluation of a core collection to identify resistance to late leafspot in peanut. Crop Sci. **35**, 1700–1702.

HUAMAN, Z., R. ORTIZ, D. ZHANG and F. RODRIGUEZ 2000: Isozyme analysis of entire and core collections of *Solanum tuberosum* subsp. *Andigena* potato cultivars. Crop Sci. **40**, 273–276.

LI, Z.C., H.L. ZHANG, Y.W. ZENG, Z.Y. YANG, S.Q. SHEN, C.Q. SUN and X.K. WANG 2002: Studies on sampling schemes for establishment of core collection of rice landraces in Yunnan, China. Genet. Resour. Crop Evol. 49, 67–74.

LI, Z.C., H.L. ZHANG, Y.S. CAO, Z.E. QIU, X.H. WEI, S.X. TANG, P. YU and X.K. WANG 2003: Studies on the sampling strategy for primary core collection of Chinese indigenous rice. Acta Agr. Sinica. **29**, 20–24.

MALOSETTI, M. and T. ABADIE 2001: Sampling strategy to develop a core collection of Uruguayan maize landraces based on morphological characters. Genet. Resour. Crop Evol. **48**, 381–390.

ORTIZ, R., E.N. RUIZ-TAPIA and A. MUJICA-SANCHEZ 1998: Sampling strategy for a core collection of Peruvian quinoa germplasm. Theor. Appl. Genet. **96**, 485–493.

PERRY, M.C., M.S. MCINTOSH and A.K. STONER 1991: Geographical patterns of variation in the USDA soybean germplasm collection: Allozyme frequencies. Crop Sci. **31**, 1356–1360.

SPAGONLETTI, Z.P.L. and C.O. QUASLSET 1991: Evaluation of five strategies for obtaining a core subset from a large genetic resources collection of durum wheat. Theor. Appl. Genet. **87**, 731–736.

SUN, C.Q., Z.C. LI and X.K. WANG 2001: Studies on evaluation of the genetic diversity of core collection of common wild rice (*Oryza rufipogon* Griff.) and Asian cultivated rice (*Oryza stativa* L.). Acta Agr. Sinica. **27**, 313–318.

UPADHYAYA, H.D., P.J. BRAMEL, R. ORTIZ and S. SINGH 2002: Developing a mini core of peanut for utilization of genetic resources. Crop Sci. **42**, 2150–2156.

UPADHYAYA, H.D., R. ORTIZRTIZ, P.J. BRAMEL and S. SINGH 2003: Development of a groundnut core collection using taxonomical, geographical and morphological descriptors. Genet. Resour. Crop Evol. **50**, 139–148.

WANG, L.X., Y.H. LI, W. LI, L. ZHU, Y. GUAN, X.C. NING, R.X GUAN, Z.X. LIU, R.Z. CHANG and L.J. QIU 2004: Establishment of a core collection of Changjiang spring sowing soybean. Chinese Biodiversity **12**, 578–585.

WANG, Z.H. and Z.X. LU 1990: Study on the evolution and classification of peach varieties through isozyme analysis. Acta Hortic. Sinica. **17**, 241–248.

WANG, Z.H. and E.J. ZHUANG 2001: China Fruit Flora. Peach. Chinese Forest Press, Beijing, 9–51.

ZHANG, H.L., Z.C. LI, Y.S. CAO, Z.E. QIU, P. YU and X.K. WANG 2003: Comparison of parameters for testing the rice core collection in phenotype. Acta Agr. Sincia. **29**, 252–257.

Received May 19, 2006 / Accepted December 23, 2006

Addresses of authors: Y. X. Li and T. H. Li (corresponding author) , Department of Fruit Science, College of Agronomy and Biotechnology, China Agricultural University, 100094 Beijing, and H. L. Zhang and Y. W Qi, College of Crop Science, China Agricultural University, Beijing, P. R. China,
e-mail: lth123430@sohu.com.