Core-Germplasm Construction of *Rosa rugosa* Collections by CDDP Molecular Markers

Liyuan Jiang, Dekui Zang*

Forestry College, Shandong Agricultural University, Tai'an, China
Email: *zangdk@sdau.edu.cn

Abstract

A method for constructing core-germplasm of *Rosa rugosa* in China based on molecular marker data was probed and the optimum core germplasm was established. Studies were initiated to analyze the genetic diversity of 120 rugged roses from 6 different Chinese source populations based on CDDP marker, and a preliminary construction of the core collection was established using stepwise UPGMA clustering sampling method. The 26 core collection resources of *R. rugosa* collections in China have 20% germplasm samples of initial collection, the retention ratio of polymorphic loci, effective number of alleles (Ne), Nei's genetic diversity (H) and Shannon information index (I) were respectively 97.52%, 104.16%, 108.38% and 106.18%. The results of t-test showed that no significant difference was found in genetic diversity indexes between the core collection and the original collection. These results also demonstrated that the core collection could stand for original collection excellently. The results show that CDDP molecular technology can be successfully applied to the construction of core germplasm resources of rugged roses.

1. Introduction

Rosa rugosa, a deciduous shrub of the genus Rosa, has prickly stems, pinnately compound leaves, and variously colored, often fragrant flowers. The rugged rose is a deciduous shrub native to northeastern China and adjacent areas of North Korea, Japan and eastern Russia [1] [2]. In China, the rugged rose is naturally distributed on the coast and islands of Southern Liaoning Province, Eastern Shandong Province, and Tumen River estuary in Jilin Province, and is classified as an endangered species [1]. *R. rugosa* is economically important in many fields

because their petals are a source of rose essential oil, which is also known as "liquid gold", a very valuable natural raw material, especially in perfumes, cosmetics, aromatherapy, spices, and nutrition [3]. As *R. rugosa* is harder than most cultivated roses, its germplasm is an important resource for rose cultivation. This species is also important ecologically, as it can grow on sandy substrates and thus protect delicate coastal areas from erosion. However, in recent years, many of the coastal areas that comprise the endemic habitat of the rugged rose have been subject to increasing industrial development, tourism, aquaculture, and other detrimental human activities, resulting in a decrease in suitable habitats [4]. In order to protect endemic populations of this ecologically important plant in China better, it is necessary to have a deep understanding of genetic variation and establish a core collection among various rugged rose populations. Collard and Mackill developed a new molecular marker technique named CDDP (conserved DNA-derived polymorphism) molecular marker, which is an effective supplement to labeling methods such as RAPD, ISSR, TRAP, CoRAP, and SCoT, and is prospectively useful in many applications [5]. In 1984 Frankel proposed [6] [7] the core collection concept, which means it is possible to select some of the numbers and the minimum genetic duplication of resources from the entire genetic resources, to maximize the diversity, preservation and utilization of these sources on behalf of the entire sample, so as to facilitate the evaluation of germplasm resources. Frankel and Brown will be further developed into a new approach for the construction of germplasm resources in-depth evaluation as well as effective protection and utilization, and gradually into a hot spot of international plant germplasm resource research [8]. The core collection of many crops has been established, such as *Triticum aestivum* [9], *Glycine max* [10] and *Medicagosativa* [11]. At present, molecular markers have been applied to the establishment of core collections of *Malus sieversii* [12] [13], *Prunus persica* [14], *Ginkgo biloba* [15] and many other species. The aim of this study is to establish core germplasm resources of rugged rose with CDDP molecular marker technology.

2. Materials and Methods

2.1. Plant Materials

We collected fresh leaves from 120 different *R. rugosa* plants in various locations in northeast China (Table 1). All specimens were preserved in silica gel at −20˚C.

2.2. DNA Extraction and CDDP Marker Analysis

We extracted the total genomic DNA of each plant from the leaves (n = 120) using the cetyl trimethyl ammonium bromide (CTAB) method (Murray and Thompson, 1980) with minor modifications. We measured DNA quantity with a spectrophotometer and DNA quality using agarose gel electrophoresis. All DNA samples were stored at −20˚C before use. We tested 21 CDDP primers (Collard *et al.* 2009; synthesized by Shanghai Biotechnology Engineering Co., Ltd).

Table 1. The120 *R. rugosa* samples used in this study.

ID Code	Samplesite	Latitude	Longitude	Elevation (meters)
1 - 45	Huichun, Jilin Province (A)	42.36˚	130.31˚	5.5
46 - 60	Mingyang Town, Liaoning Province (B)	39.33˚	122.40˚	3.0
61 - 75	Muping District, Yantai City (C)	37.46˚	121.69˚	4.5
76 - 90	Jianggezhuang Town, Yantai City (D)	37.45˚	121.79˚	5.0
91 - 105	Shichang Road, Weihai City (E)	37.50˚	122.00˚	3.0
106 - 120	Chengshan Town, Weihai City (F)	37.40˚	122.54˚	3.0

Table 2. 13 primers for rugged rose CDDP molecular marker.

Code	Aminoacidmotif	Primername	Primersequence (5'-3')	Length	GC (%)
Pr1	MGRGKV	MADS-2	ATGGGCCGSGGCAAGGTGG	19	74
Pr2	TPIHR	ABP1-1	ACSCCSATCCACCGC	15	73
Pr3	HYRGVR	ERF1	CACTACCGCGGSCTSCG	17	77
Pr4	MGRGKV	MADS-1	ATGGGCCGSGGCAAGGTGC	19	74
Pr5	GKSCR	Myb1	GGCAAGGGCTGCCGC	15	80
Pr6	AEIRDP	ERF2	GCSGAGATCCGSGACCC	17	77
Pr7	GEHTC	WRKY-R3B	CCGCTCGTGTGSACG	15	73
Pr8	GKHNH	WRKY-R1	GTGGTTGTGCTTGCC	15	60
Pr9	GEHTC	WRKY-R3	GCASGTGTGCTCGCC	15	73
Pr10	HWWELH	KNOX-2	CACTGGTGGGAGCTSCAC	18	67
Pr11	WRKYGQ	WRKY-F1	TGGCGSAAGTACGGCCAG	18	67
Pr12	TPIHR	ABP1-2	ACSCCSATCCACCGC	15	73
Pr13	GKSCR	Myb2	GGCAAGGGCTGCCGG	15	80

Of these, preliminary results indicated that 13 resulted in clear, stable and polymorphic bands. These 13 primers (Table 2) were selected for PCR amplification. Each 20 μL PCR mixture contained 10 μL 2XTaq Master Mix (containing dye), 1 μL 40 ng/μL DNA template, 1.0 μL 10 pmol/μL primer, and 8 μL ddH$_2$O. A standard PCR cycle was used: an initial denaturation step at 94˚C for 3 min; 35 annealing cycles of 94˚C for 1 min, 50˚C for 1 min and 72˚C for 2 min; a final extension of 5 min at 72˚C. PCR products were stored at 4˚C. PCR products were separated by electrophores is on 2% polyacrylamide gel using a vertical-gel apparatus in 1 × TAE at 120V constant current for 1.5 - 2 h, along with a DL5000 marker as a size marker. Amplified fragments were strained by the silver-straining method and photographed with a digital camera. All amplifications were repeated at least twice.

2.3. Data Scoring and Statistical Analysis

Based on the electrophorogram, we selected only clear bands for statistical analysis. Only bands that were clearly polymorphic in all replicates were scored as

present (1) in a data matrix. Bands without clear polymorphisms, or that were only polymorphic in one replicate were scored as absent (0). We calculated the total number of bands, the number of polymorphic bands, and the percentage of the total bands that were polymorphic or specific using Microsoft Excel 2007. We used POPGENE 1.32 to compute: the number of effective loci, the percentage of polymorphic loci, Shannon's information index (I), observed number of alleles (Na), expected heterozygosity (He), effective number of alleles (Ne), Nei's gene diversity (H). We constructed a UPGMA dendrogram using SHAN. We used Tree Plot to draw a cluster analysis tree (Xu 1994).

The calculation methods of the above parameters are as follows:

Effective number of alleles (Ne) = $1/\sum Pi2$, Pi is the frequency of variation in the first i allele; Nei's gene diversity (H) = $1 - \sum Pik2$, Pik is the frequency of the occurrence of i alleles in all genes; Shannon's information index (I) = $-\sum (Pi \times \ln Pi)$, Pi is the frequency of variation in the first i allele.

2.4. Construction and Evaluation of Core Collection

NTSYS software was adopted to cluster the experimental materials and all the germplasms were grouped with the appropriate genetic similarity coefficients. In the whole sampling, the number of each sample accepted into the core collection was determined according to the sampling ratio given. They take different proportions (80%, 70%, 60%, 50%, 40%, 30%, 20%, 10%). Then the genetic similarity coefficient of the germplasm of pairs from each group was selected and the poor ones were deleted. The process went on and eventually the representative germplasms that meet the requirements constituted a core subset.

The genetic diversity parameters of the original germplasm and every candidate core collection were carried out with T test with SPSS statistical software. The significance of T value was used to represent the significant difference between the candidate core germplasm and the original germplasm. Based on this, whether the core collection is reasonable can be further determined.

3. Results

3.1. Genetic Diversity Analysis of *R. rugosa* Population

Among the 128 fragments generated from the 13 selected CDDP primers, 121 were polymorphic (94.53%), an average of 9.85 polymorphic bands perprimer. This suggests that rugged roses are rich in genetic diversity. At population level, the genetic diversity index is the highest in Hunchun (A), Chengshan (F) minimum (Table 3). Therefore, the genetic diversity of Hunchun population in Northeast China is the highest.

3.2. Construction of Wild Rose Core Collection

The UPGMA dendrogram showed that the populations of *R. rugosa* analyzed could be divided into six major groups when genetic coefficient is 0.81 (Figure 1). The core collection was screened according to these groups.

Table 3. Genetic diversity of the six populations of *R. rugosa.*

Population	Number of polymorphic loci(NPL)	Percentage of polymorphism (PPL)	Effective number of alleles(Na)	Effective number of alleles (Ne)	Nei's gene diversity index (H)	Shannon's information index (I)
Hui Chun (A)	88	68.75%	1.8125	1.4159	0.2548	0.3861
Ming Yang (B)	30	23.44%	1.2344	1.1682	0.0961	0.1400
Mou Ping (C)	51	39.84%	1.4062	1.2818	0.1601	0.2339
Yun Xi (D)	56	43.75%	1.4297	1.2933	0.1640	0.2400
Shi Chang (E)	57	44.53%	1.4453	1.2863	0.1690	0.2505
Cheng Shan (F)	22	17.19%	1.1719	1.1239	0.0695	0.1011

The results of stepwise cluster analysis showed that the number of polymorphic loci (NPL) decreased gradually with the decrease of sampling proportion, while the values of Ne, H and I increased gradually (Table 4). If the primary core collection of

R. rugosa is constructed, the core collection needs less resources to represent more than 95% of the genetic diversity of the original germplasm. Generally it is considered that 20% - 30% of original germplasm as core germplasm is the best [16]. Overwhelmingly, a sampling ratio of 20% was selected as core germplasm. The codes are: C1, C7, C15, C19, C21, C25, C31, C32, C42, C44, C46, C50, C58, C60, C61, C63, C68, C76, C85, C88, C91, C93, C103, C104, C107, C117.

3.3. Evaluation of Core Collection

The genetic similarity coefficient (GS) for all 26 core samples ranged from 0.67 to 0.97 (Figure 2). Compared with the original germplasm (0.71 - 1.00), the relationship between the germplasm becomes far away. The 26 core collection resources of *R. rugosa* collections in China have 20% germplasm samples of initial collection, the retention ratio of polymorphic loci, effective number of alleles (Ne), Nei's genetic diversity (H) and Shannon information index (I) were respectively 97.52%, 104.16%, 108.38% and 106.18%. The genetic parameters of selected core germplasms was obtained with T test and the results showed that it was not distinct from those of the original germplasm (Table 5).

4. Discussion

4.1. Method for Constructing Core Collection

The selection of sampling methods has always been the focus of the construction of core germplasm because it determines which germplasm is eligible for core collection. According to the objective of core collection, core germplasm should represent the greatest genetic diversity of the entire genetic resource. In this study, the core collection of wild roses was constructed with the method of stepwise clustering. This method has been used by many researchers [17] [18] [19].

Establishment of core collection is mostly based on morphological traits, but

Figure 1. UPGMA dendrogram of genetic similarity of 120 samples among rugged rose.
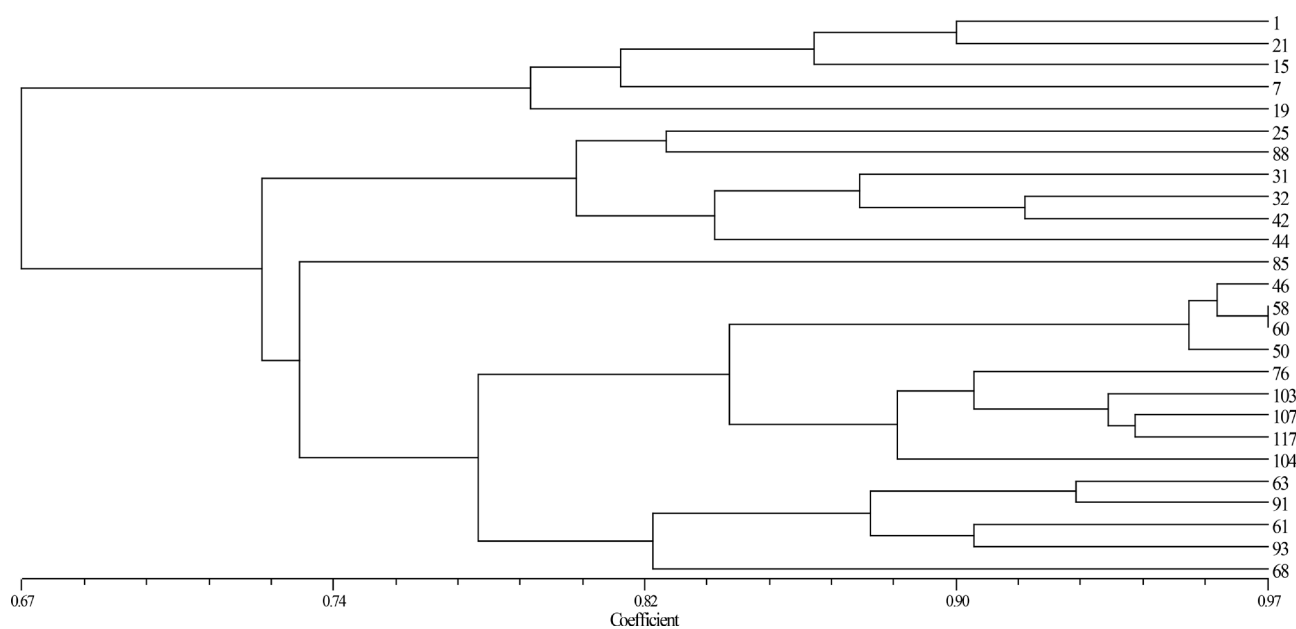
Figure 2. UPGMA dendrogram of genetic similarity of 26 core samples among rugged rose.

Table 4. Genetic diversities of core germplasm constructed by the grouping gradual clustering method.

Sampling ratio	Number of samples	NPL/%	Na	Ne	H	I
100%	120	121	1.9453	1.5149	0.3089	0.4691
80%	96	120 (99.17%)	1.9375	1.5329	0.3162	0.4770
70%	85	120 (99.17%)	1.9375	1.5420	0.3200	0.4815
60%	73	119 (98.35%)	1.9297	1.5585	0.3268	0.4892
50%	61	119 (98.35%)	1.9297	1.5610	0.3280	0.4907
40%	49	118 (97.52%)	1.9219	1.5655	0.3291	0.4909
30%	37	118 (97.52%)	1.9219	1.5723	0.3319	0.4943
20%	26	118 (97.52%)	1.9219	1.5778	0.3348	0.4981
10%	13	113 (93.39%)	1.8828	1.6105	0.3422	0.5013

Table 5. Comparision of the genetic diversities between the core collection and the primary group of *R. rugosa*.

Genetic diversities		Primary group	Core germplasm	$|t|$ value	Percent of reserve/%
Ne	Value	1.5149	1.5779	1.560	104.16
	Var	0.100904	0.107439		
H	Value	0.3089	0.3348	1.351	108.38
	Var	0.023006	0.024082		
I	Value	0.4691	0.4981	1.160	106.18
	Var	0.038615	0.041333		

$a = 0.05$, $t_a = 1.960$

because of the morphological characters of vulnerable ecological environment of plant growth and various climate factors, full use of agronomic or morphological data sampling to construct nuclear germplasm and genetic resources is not representative for all species [18] [19]. However, molecular markers can exclude genetic differences in environmental traits [20]. There are a variety of molecular screening technology is that have been applied to core collection. ISS molecular marker technique was successfully applied to the core germplasm screening of *Morinda citrifolia* [21]. 35% of the core collection of *Ginkgo biloba was* retained *u*sing APLF markers [15]. The core collections in the French Apple germplasm based on SSR markers was constructed [22]. In this study, we adopted CDDP molecular marker technology, which is rich in polymorphism and information and high in versatility [23]. The results showed that 20% core collection of *R. rugosa* was screened out by using CDDP molecular technique.

4.2. Reasonable Sampling Proportion

The purpose of core collection is to obtain the diversity that represents the entire genetic resource to the greatest extent with a minimum number. Therefore, the rational sampling proportion is the key point. Brown believes that the core collection samples from 5% to 10% of the original germplasm resources can represent more than 70% of the genetic variations of the whole germplasm [8]. In core collection construction, the proportion of core germplasm generally accounts for 5% ~ 40% of the original Germplasm [24] [25]. Li believes the sampling ratio should be determined based on the genetic structure and the number of specific species. The proportion of core germplasm in the small germplasm of the original germplasm is relatively large, whereas the proportion of core germplasm is smaller [26]. Zhang screened 30 samples to construct the core collection of 100 cultivars of *osmanthus fragrans* [27]; Liu selected 25 samples from 110 *pummelo* germplasm resources to construct core germplasm of pomelo [28]. Because there are only 120 copies in the original germplasm in this study, in order to minimize the loss of genetic variation and to ensure that the sampling points can be built into the core germplasm collection, finally a collection contains 26 core germplasms of different population groups, accounting for 20% of the proportion of the original germplasm sampling, is obtained. The results show that the diversity and genetic diversity structure of the original population can be well preserved. In addition, according to the result of T test, the genetic diversity of the core collection is not significantly different from that of the original germplasm. Since all rugged rose resources are not included in this study, the core collection needs to be further supplemented with new germplasm resources to ensure the long-term conservation and utilization of the diversity and resources of the species.

4.3. Conservation Strategies for *R. rugosa*

With the development of industry, the population size of the rugged rose is de-

creasing. As a result, it is urgent to develop reasonable strategies for the conservation and breeding of germplasm resources. Not only the large number but also the repetition rate up to 60% of the collections has become the major obstacle for effective use of germplasm materials collected in most plant resources [29]. Therefore, we propose the following approach. First, we should focus on areas that are rich in genetic diversity (Hunchun, Jilin Province). Second, the rugged rose core germplasm resources selected from this study should be sent to the germplasm database for protection and research.

5. Conclusion

In this study, based on CDDP molecular markers, the core germplasm resources of rugged rose in China were constructed with the method of stepwise clustering. Core collection can provide a theoretical basis for the protection and utilization of wild rose germplasm resources. At the same time, it provides a reference for the application of CDDP molecular technology for other plants.

Acknowledgements

References

[1] Fu, L.G. and Jin, J.M. (1992) China Plant Red Data Book Rare and Endangered Plants. Science Press, Vol. 1, 558-559.

[2] Bruun, H.H. (2005) *Rosa rugosa* Thunb.ex Murray. *Journal of Ecology*, **93**, 441-470. https://doi.org/10.1111/j.1365-2745.2005.01002.x

[3] Ma, X.H., Wang, Y.H., Wei, Q. and Zhang, G.J. (2004) Study on Processing Technology of Rose Essential Oil. *Chemistry and Industry of Forest Products*, **24**, 80-84.

[4] Tong, R., Wu, X.L., Jiang, L.N., Si, Q.Q. and Zang, D.K. (2017) Phenotypic Variations in Populations of *Rosa rugosa*. *Acta Ecologica Sinica*, **37**, 3706-3715.

[5] Collard, B.C.Y. and Mackill, D.J. (2009) Conserved DNA-Derived Polymorphism (CDDP): A Simple and Novel Method for Generating DNA Markers in Plants. *Plant Molecular Biology Reporter*, **27**, 558-562.
https://doi.org/10.1007/s11105-009-0118-z

[6] Frankel, O.H. (1984) Genetic Perspectives of Germplasm Conservation. In: Arber, W., Illmensee, K., Peacock, W.J. and Starlinger, P., Eds., *Genetic Manipulation: Impact on Man and Society*, Cambridge University Press, Cambridge, England, 161-170.

[7] Frankel, O.H. and Brown, A.H.D. (1984) Plant Genetic Resources Today: A Critical Appraisal. In: Holden, J.H. and Williams, J.T., Eds., *Crop Genetic Resources: Conservation and Evaluation,* George Allen and Unwin, London, 249-257.

[8] Brown, A.H.D., Frankel, O.H. and Marshall, R.D. (1989) The Use of Plant Genetic Resources. Cambridge University Press, Cambridge, England, 136-156.

[9] SpagnolettiZeuli, P.L. and Qualset, C.O. (1993) Evaluation of Five Strategies for Obtaining a Core Subset from a Large Genetic Resource Collection of Durum

Wheat. *Theoretical and Applied Genetics*, **87**, 295-304.
https://doi.org/10.1007/BF01184915

[10] Qiu, L.J., Cao, Y.S., Chang, R.Z., Zhou, X.A., Wang, G.X., Sun, J.Y., Xie, H., Zhang, B., Li, X.H., Xu, Z.Y. and Liu, L.H. (2003) Establishment of Chinese Soybean (G.max) Core Collection Sampling Strategy. *Scientia Agricultura Sinica*, **36**, 1442-1449.

[11] Diwan, N., Mcintosh, M.S. and Bauchan, G.R. (1995) Methods of Developing a Core Collection of Annual Medicago Species. *Theoretical and Applied Genetics*, **90**, 755-761. https://doi.org/10.1007/BF00222008

[12] Zhang, C.Y. (2008) Population Genetic Structure and Method of Constructing Core Collection for *Malus sieversii*. Shandong Agricultural University.

[13] Zhang, C.Y., Chen, X.S., Zhang, Y.M., Yuan, Z.H., Liu, Z.C., Wang, Y.L. and Lin, Q. (2009) A Method for Constructing Core Collection of *Malus sieversii* using Molecular Markers. *Scientia Agricultura Sinica*, **42**, 597-604.

[14] Li, Y.X., Gao, Q.J. and Li, T.H. (2006) Sampling Strategy Based on Fruit Characteristics for a Primary Core Collection of Peach Cultivars. *Journal of Fruit Scienca*, **23**, 359-364.

[15] Wang, X., Liu, X.J. and Xing, S.Y. (2009) AFLP Analysis of Genetic Diversity and a Construction of the Core Collection of Partial Ancient Ginkgo Trees in China. *Acta Horticulturae Sinica*, **43**, 249-260.

[16] Liu, Z.C., Zhang, C.Y., Zhang, Y.M. and Zhang, X.Y. (2010) Study on Method of Constructing Core Collection of *Malus sieversii* Based on Quantitative Traits. *Scientia Agricultura Sinica*, **43**, 358-370.

[17] Wang, H.X., Zhao, S.G., Gao, Y., Xuan, L.C. and Zhang, Z.C. (2013) A Construction of the Core-Collection of *Juglans regia* L. Based on AFLP Molecular Markers. *Scientia Agricultura Sinica*, **46**, 4985-4995.

[18] AIbrecht, C., Russinova, E. and Hecht, V. (2005) The *Arabidopsis thaliana Somatic embryogenesis* Receptor-Like Kinases 1 and 2 Control Male Sporogenesis. *Plant Cell*, **17**, 3337-3349. https://doi.org/10.1105/tpc.105.036814

[19] Colcombet, J., Boisson-Demier, A. and Ros-Palau, R. (2005) Arabidopsis Somatic Embryogenesis Receptor-Like Kinases 1 and 2 Are Essential for Tapetum Development and Microspore Maturation. *Plant Cell*, **17**, 3350-3361.
https://doi.org/10.1105/tpc.105.036731

[20] Karp, A., Kresovich, S. and Bhat, K.V. (1997) Molecular Tools in Plant Genetic Resources Conservation; A Guide to the Technologies. IPGRI Technical Bulletin No. 2. International Plant Genetic Resources Institute, Rome, 97, 47.

[21] Liao, D. (2016) Core-Germplasm Construction of *Morinda citrifolia* Collections by ISSR Molecular Markers. Hainan University.

[22] Ludivine, L., Caroline, D. and Elisa, R.D. (2016) Genetic Diversity, Population Structure, Parentage Analysis, and Construction of Core Collections in the French Apple Germplasm Based on SSR Markers Plant. *Molecular Biology Reports*, **34**, 827-844. https://doi.org/10.1007/s11105-015-0966-7

[23] Li, Y.Y. and Zheng, C.S. (2013) Genetic Diversity of Tree Peony Cultivar Resources in Heze Revealed by CDDP Marker Scientia. *Agricultura Sinica*, **46**, 2739-2750.

[24] Ortiz, R., Ruiz-Tapia, E.N. and Mujica-Sanchez, A. (1998) Sampling Strategy for a Core Collection of *Peruvian quinoa* Germplasm. *Theoretical and Applied Genetics*, **96**, 475-483. https://doi.org/10.1007/s001220050764

[25] Li, G.S., Zhang, L.J. and Bai, C.K. (2012) Chinese *Cornus officinalis*: Genetic Re-

sources, Genetic Diversity and Core Collection. *Genetic Resources and Crop Evolution*, **59**, 1659-1671. https://doi.org/10.1007/s10722-011-9789-z

[26] Li, Z.C., Zhang, H.L., Cao, Y.S., Qiu, Z.E., Wei, X.H., Tang, S.X., Yu, P. and Wang, X.K. (2003) Studies on the Sampling Strategy for Primary Core Collection of Chinese Ingenious Rice. *Scientia Agricultura Sinica*, **29**, 20-24.

[27] Zhang, W.R., Yuan, W.J. and Shang, F.D. (2012) Development of Core Collection of Osmanthus Fragrans Lour. Cultivars Based on AFLP Molecular Markers. *Acta Botanica Boreali-Occidentalia Sinica*, **32**, 1349-1354.

[28] Liu, Y., Sun, Z.H., Liu, D., Wu, B. and Zhou, Q. (2006) Screening the Core Collection of Pomelo Germplasm Based on Molecular Marker. *Journal of Fruit Science*, **23**, 339-345.

[29] Zhou, M.D. (1991) Core Collection of Crop Germplasm. Newsletter of IPGRI, No. 5.