Front. Agric. China 2011, 5(3): 344–350
DOI 10.1007/s11703-011-1097-z

RESEARCH ARTICLE

Establishment of core collection for Chinese tea germplasm based on cultivated region grouping and phenotypic data

Xinchao WANG, Liang CHEN (✉), Yajun YANG (✉)

Tea Research Institute of the Chinese Academy of Agricultural Sciences; National Center for Tea Improvement, Hangzhou 310008, China

Abstract **Tea is an important economic crop in China. About 2665 accessions of tea germplasm (*Camellia* spp.) are conserved in the China National Germplasm Tea Repository. However, only a small fraction of the collections have been exhaustively used in tea plant improvement programs. The objective of the present study was to develop a core collection of tea plant to enhance the utilization of genetic resources in improvement programs and simplify their management. For this purpose, based on the cultivated region grouping and phenotypic trait data, a core collection of tea plant was established using logarithm proportion and Ward's cluster method. Approximately 20% of the accessions were then randomly selected from these distinct groups to form a core collection of 532 accessions. Different statistical methods including comparison of mean using Newman-Keuls test, variance using Levene's test, and frequency distribution using Chi-square test for the traits validated that the variation present in the initial collection was retained in the core collection. The Shannon-Weaver diversity index for different traits was also similar in the core and initial collection. The phenotypic correlations among different quantitative traits that may be under the control of coadapted gene complexes were also preserved in the core collection. This core collection will play an important role in the utilization of tea germplasms.**

Keywords tea plant (*Camellia sinensis*), germplasm, core collection, cultivated region, phenotypic trait

Introduction

Tea plant originated from the Yunnan Province of south-west China, where there are rich and diverse tea genetic resources (Yu, 1986). In the past two decades, tea genetic resources have been broadly collected, conserved, evaluated, and utilized in China. Two permanent exsitu conservation facilities, the China National Germplasm Tea Repository (CNGTR), including the China National Germplasm Hangzhou Tea Repository (CNGHTR) in the Tea Research Institute of the Chinese Academy of Agricultural Sciences (TRICAAS) and the Menghai Tea Repository Branch (CNGMTRB) in the Tea Research Institute Yunnan Academy of Agricultural Sciences (TRIYAAS), were established in 1990. Up to 2004, 2665 accessions including wild tea plants, landraces, improved cultivars, introduced varieties, genetic materials, and related species were preserved in the CNGTR (Chen et al., 2004). Agronomic traits, made tea quality, chemical components, tolerance, and resistance to biotic and abiotic stresses, of more than 1600 tea genetic resource accessions have been evaluated and appraised using multidisciplinary approaches in the past 20 years, and several elite and/or special germplasms have been identified and utilized for tea breeding (Chen et al., 2004).

Recently, following the tendency of aggravated competition on tea industry among international and domestic trades, tea industry tends to diversify in market demand. However, the shortage of elite cultivars satisfactory to the market has become an obstacle to tea industry sustainable development. Tea genetic resources are one of the most valuable and fundamental materials for breeding. Numerous tea plant germplasms have been conserved in tea repositories, and the absence of knowledge on some important traits and/or complex traits of in the genetic resources has hindered their utilization. However, the large number of accessions results in complications as their evaluation, management, and utilization are difficult, particularly in the case of initiation of new

Received January 24, 2010; accepted February 18, 2011
Correspondence: [a]Liang CHEN; [b]Yajun YANG
E-mail: [a]liangchen@mail.tricaas.com; [b]yjyang@mail.tricaas.com

breeding programs for the improvement of agronomic traits of resistance to diseases and pests as well as quality. This is because the screening of the initial collection for such traits might be quite expensive and time consuming.

Core collections, with a subset of accessions selected to represent the genetic diversity in the collection and provide a minimum of repetitiveness, have been proposed as a good way of maintaining and managing the germplasm collections (Frankel, 1984; Brown, 1989). The core collection as a convenient working collection can be evaluated and utilized preferentially, which could solve the problem in large-scale collection hindering the preservation and utilization of germplasm resources, which has received extensive attention all over the world. A great number of core collections have been established, such as rice (Li et al., 2004), maize (Li et al., 2004), wheat (Balfourier et al., 2007), etc. In tea, Li and Jiang (2004) established a core collection of 615 tea accessions based on 10 traits data. However, the development of a more representative and abundant genetic diversity core collection of tea plant in China is necessary. The objectives of our research were therefore to characterize the tea plant accessions available in CNGTR's genebank using passport data and botanical descriptors and analyze this diversity to define a core collection, which may be a valuable entry point to the initial collection for further research or use by tea plant breeders.

Materials and methods

Materials and data standardization

The CNGTR holds 2665 accessions tea germplasms, including five species and two varieties of Sect. *Thea* (L.) Dyer (Chen et al., 2000), originating from eight countries, such as China, Japan, Kenya, Georgia, India, Vietnam, Burma, Sri Lanka, etc. About 1669 accessions were systematically appraised and comprehensively evaluated using multidisciplinary approaches (Chen et al., 2004; Wang et al., 2008). Data of 33 passport data and phenotypic traits, including agronomical, cup tea quality, and main chemical components, were used to establish core collection (Table 1). The data were arranged according to procedures outlined in "Descriptors and Data Standard for Tea (*Camellia* spp.)" (Chen et al., 2005), and the qualitative traits were assigned as 1, 2, 3, etc. (Table 2). Because the number of style (1 to 5) splitting was a continuous quantitative trait, this trait was considered as a qualitative trait, and the quantitative traits were divided to 10 classes based on means $\pm 0.5\sigma$ as 1, 2, 3… 10 (Qiu et al., 2003).

Table 1 Characteristic datum used in primary core collection sampling

Data types	Character types
Passport data	Species, germplasm type, propagating type
Morphological data	Plant type, growth habit, young shoot color, young shoot pubescence, length of 'three and a bud' (LB), weight of 100 'three and a bud' (WB), leaf attitude, leaf size, leaf shape, leaf color, leaf upper surface, leaf cross section, leaf texture, sharpness of leaf serration, density of leaf serration, depth of leaf serration, leaf base shape, leaf apex shape, calyx pubescence, flower diameter (FD), ovary pubescence, style length (SL), number of style splitting, relative height between gynoecium and androecium (RHGA)
Quality characteristic data	Processing suitability, water extracts (WE), caffeine (CAF), tea polyphenols (TP), amino acids (AA), ratio of polyphenols/ amino acids (TP/AA)

Establishment of core collection

A method of stratification in cultivated ecological areas and then cluster sampling was adopted by recording the number of accessions from each cultivated ecological area, calculating the logarithm (log) of the number of accessions from each cultivated ecological area and analyzing hierarchical clusters on the standardized data using Ward's minimum variance method (Ward, 1963) with 0.75 R^2 (squared multiple correlation) for grouping the accessions. The Ward's method was used with PROC-CLUSTER program in SAS 8.0 for Windows (SAS Institute, 1989); approximately 20% accessions were randomly selected from each cluster; some germplasms with special traits, such as white young shoot and "Zigzag" branch, were directly selected into the core collection, and the germplasms were selected and not evaluated based on the logarithm (log) of the number of accessions from each cultivated ecological area (Wang et al., 2009).

Validating the core collection

The means of the core collection and the initial collection were compared using the Newman-Keuls procedure (Newman, 1939; Keuls, 1952) for all the nine quantitative traits. The homogeneity of variances of the core and initial collections was tested using Levene's test (Levene, 1960). The percentage of significant differences between the core collection and initial collection was calculated for the mean difference percentage (MD%) or the variance difference percentage (VD%) (Hu et al., 2000). The coincidence rate (CR%) and variable rate (VR%) were calculated to evaluate properties of the core collection (Hu et al., 2000). The homogeneity of distribution of the initial and core collection for each trait was measured by comparing the frequency distribution of the trait using the chi-square test. Phenotypic correlations between the nine quantitative traits in both the core and initial collections were estimated independently to assess whether these associations under same genetic control, which were conserved in the core collection. The diversity index (H') of Shannon–Weaver (1949) was used as a measure of the phenotypic diversity of each trait. The index was

Table 2 Code designed for qualitative traits in tea germplasms (*Camellia* spp.)

Character type	Code
Species	1: *C. sinensis* var. *sinensis*, 2: *C. sinensis* var. *assamica*, 3: *C. sinensis* var. *pubilimba*, 4: *C. gymnogyna*, 5: *C. taliensis*, 6: *C. crassicolumna*, 7: *C. tachangensis*, 8: *Camellia* sp.
Germplasm type	1: Wild, 2: Landrace, 3: Improved cultivar, 4: Breeding line, 5: Genetic stock, 6: Other
Propagating type	1: Sexual, 2: Asexual
Plant type	1: Shrub, 2: Semiarbor, 3: Arbor
Growth habit	1: Erect, 2: Semi-erect, 3: Horizontal spreading
Young shoot color	1: Whitish, 2: Yellow green, 3: Light green, 4: Green, 5: Purple green
Young shoot pubescence	0: Absent, 1: Sparse, 2: Medium, 3: Dense, 4: Extremely dense
Leaf attitude	1: Erect, 2: Semi-erect, 3: Horizontal, 4: Drooping
Leaf size	1: Small, 2: Medium, 3: Large, 4: Extremely large
Leaf shape	1: Near round, 2: Ovate, 3: Elliptic, 4: Oblong, 5: Lanceolate
Leaf color	1: Yellow green, 2: Light green, 3: Green, 4: Dark green
Leaf upper surface	1: Smooth, 2: Slightly rugose, 3: Rugose
Leaf cross section	1: Convex, 2: Flat, 3: Concave
Leaf texture	1: Soft, 2: Medium, 3: Hard
Sharpness of leaf serration	1: Sharp, 2: Medium, 3: Obtuse
Density of leaf serration	1: Sparse, 2: Medium, 3: Dense
Depth of leaf serration	1: Flat, 2: Medium, 3: Deep
Leaf base shape	1: Acute, 2: Obtuse
Leaf apex shape	1: Acute, 2: Attenuate, 3: Blunt, 4: Obtuse
Calyx pubescence	0: Absent, 1: Present
Ovary pubescence	0: Absent, 1: Present
Relative height between Gynoecium and androecium	1: Gynoecium lower, 2: Gynoecium and androecium same height, 3: Gynoecium higher
Processing suitability	1: Green tea, 2: Black tea, 3: Oolong tea, 4: Unsuitable

calculated independently in both the core collection and initial collection to determine whether the diversity was retained in the core collection. The coincidence rate of index numbers (CRIN%) was used to test whether the core collection retained enough variability (Zhao et al., 2005). Data analysis was done using SAS software (SAS Institute, 1989).

$H' = -\sum P_j \ln P_j$, where P_j is the traits appearing frequency of the code j.

Coincidence rate of index numbers $(\mathrm{CRIN}\%) = \sum (N_j' / N_j) \times 100$, where N_j' is the traits indicator of core collection of number j, and N_j is the initial germplasm traits indicator of the number j.

Results

Core collection

In accordance with the national tea production zoning, the tea cultivated regions in China are divided into four: North of the Yangtze River (NYR), South of the Yangtze River (SYR), South China (SC), and South-west China (SWC) (Tea Research Institute Chinese Academy of Agricultural Sciences, 1986). The introduced genetic resources are dealt with as a separate group; hence, there are five groups. The grouping result of each tea production region is shown in Table 3. Using the logarithmic sampling strategy, as per the proportion of 20% of the overall sample, the primary core collection of tea plant contained 532 accessions, among which there were 79 accessions from NYR, 125 accessions from SYR, 128 accessions from SC, 124 accessions from SWC, and 76 accessions from the introduced resources. Compared with the initial collection, in the core collection, the proportion of NYR (from 4.7% to 14.8%) and the introduced resources (from 4.0% to 14.3%) increased, the proportion of SYR (from 29.8% to 23.5%) and SC (from 34.4% to 24.1%) reduced, and the ratio of SWC had a slight change (Table 3).

Validating the core collection sampling

Differences between the means of the initial and core collections were found to be nonsignificant for all traits except for CAF (Table 4). The recorded MD% (11.1%) indicated that the core collection represented the initial collection adequately (Hu et al., 2000). The variances of the

Table 3 Grouping results of tea germplasms in the initial and core collections of China

Cultivated region	Initial collection			Core collection		
	Number	Percentage (%)	Proportion logarithm	Number	Percentage of initial collection	Percentage of core collection
NYR	125	4.7	14.9	79	3.0	14.8
SYR	795	29.8	23.5	125	4.7	23.5
SC	916	34.4	24.1	128	4.8	24.1
SWC	721	27.1	23.2	124	4.6	23.3
Introduction	108	4.0	14.3	76	2.9	14.3
Total	2665	100.0	100.0	532	20.0	100.0

Table 4 Comparison of mean, range, coefficient of variation, and variance for nine quantitative traits in the initial and core collections of tea plant

Trait	Initial collection				Core collection					
	Mean±*SD*	Range	CV (%)	Variance	Mean±*SD*	Range	CV (%)	Variance	Levene's-test	*F*-test
LB	7.3±1.8	3.1–15.2	24.1	3.14	7.3±1.9	3.4–13.2	25.5	3.46	NS#	NS
WB	62.7±29.8	16.5–205.0	47.5	891.53	63.4±30.2	22.3–171.9	47.6	909.97	NS	NS
FD	3.7±0.6	1.4–6.9	15.6	0.34	3.8±0.7	1.8–6.2	18.1	0.47	NS	NS
SL	1.3±0.2	0.6–2.4	17.1	0.05	1.3±0.3	0.6–2.1	19.2	0.06	NS	NS
WE	42.7±3S.9	20.9–56.1	9.0	14.87	42.5±4.5	20.9–56.1	10.6	20.17	NS	NS
CAF	4.0±0.6	0.4–6.2	16.2	0.41	3.8±0.8	0.4–6.2	21.4	0.67	*	NS
TP	28.1±4.8	12.2–41.5	17.2	23.45	27.6±5.7	13.8–40.0	20.8	32.82	NS	NS
AA	3.0±0.9	0.5–6.5	28.9	0.77	3.0±1.1	0.5–6.5	36.0	1.15	NS	NS
TP/AA	10.4±5.1	2.2–80.0	48.9	25.97	11.3±7.8	2.2–80.0	69.0	60.36	NS	NS

LB: Length of 'three and a bud', WB: Weight of 100 'three and a bud', FD: Flower diameter, SL: Style length, RHGA: Relative height between gynoecium and androecium, WE: Water extracts, CAF: Caffeine, TP: Tea polyphenols, AA: Amino acids, TP/AA: Ratio of polyphenols/ amino acids (the same as Table7). #: NS, Nonsignificant at $P = 0.05$, *: significant at $P = 0.05$.

initial and the core collections were homogeneous for all traits (Table 4). The CR% of WE, CAF, TP, AA, and TP/AA were over 89%, and the CR% of LB, WB, FD, and SL were between 75.5% and 83.3%. The high CR% retained in the core collection (89.9%) indicated that it was representative of the initial collection. The coefficient of variations for all nine traits of core collection was higher than that of the initial collection, which showed the variation of core collection was greater, and most of the traits of extreme materials were retained in the core collection (VR% = 119.0%).

The analysis of frequency distribution, except germplasm types ($P = 0.028$), confirmed the homogeneity of the distribution between the initial and collections (Table 5). Analysis of the frequency distribution of germplasm types, in the core collection, improved the share of wild germplasms, improved cultivars and related species (other), and reduced the proportion of landraces and strains. Wild germplasms and related species were likely to hold the specific traits that were not possessed in cultivation type. Moreover, the improved cultivars were vastly spread in the production, since they contained many traits suitable for cultivation. Therefore, the rising proportion of these germplasms in the ratio of core collection will help to keep some specific traits needed for production and can be used in the future for cultivar improvement.

The Shannon-Weaver diversity index (H') was calculated and compared with the phenotypic diversity among traits in the initial and core collections. The index provided a measurement of both allelic richness and allelic evenness. A low index showed that frequency classes were unbalanced for a trait lacking genetic diversity. The average indexes for the 31 traits in the core collection ($H' = 1.224$) and the initial collection ($H' = 1.184$) were similar (Table 6), indicating that the diversity of initial collection was well represented in the core collection. In addition, it can be seen that the average genetic diversity index of core collection was greater than that

Table 5 Comparison of frequency distribution of 33 traits between core and initial collection by χ^2 test

Trait	Class	χ^2	*P* value	Trait	Class	χ^2	*P* value
Species	8	8.785	0.268	Density of leaf serration	3	4.403	0.111
Germplasm type	6	10.913*	0.028	Depth of leaf serration	3	0.892	0.640
Propagating type	2	0.108	0.663	Leaf base shape	2	0.441	0.402
Plant type	3	0.090	0.956	Leaf apex shape	4	1.562	0.668
Growth habit	3	0.732	0.693	Calyx pubescence	2	0.002	0.944
Young shoot color	5	3.215	0.523	Flower diameter	10	16.492	0.183
Young shoot pubescence	5	7.844	0.098	Ovary pubescence	2	0.032	0.964
Length of 'three and a bud'	10	2.572	0.973	Style length	10	7.030	0.653
Weight of 100 'three and a bud'	10	6.039	0.756	Number of style splitting	5	2303	0.680
Leaf attitude	4	6.669	0.083	Relative height between gynoecium and androecium	3	0.326	0.850
Leaf size	4	0.759	0.859	Processing suitability	4	0.454	0.141
Leaf shape	5	2.468	0.650	Water extracts	10	14.444	0.319
Leaf color	4	0.825	0.844	Caffeine	10	22.946	0.090
Leaf upper surface	3	4.755	0.093	Tea polyphenols	10	11.930	0.370
Leaf cross section	3	1.033	0.597	Amino acids	10	18.984	0.134
Leaf texture	3	1.950	0.377	Ratio of polyphenols/amino acids	10	11.926	0.338
Sharpness of leaf serration	3	0.862	0.650				

Table 6 Comparison of Shannon-weaver diversity indices for 31 traits between the initial and core collections of tea plant

Trait	Initial collection	Core collection	Trait	Initial collection	Core collection
Propagating type	0.613	0.628	Depth of leaf serration	1.017	1.021
Plant type	0.942	0.932	Leaf base shape	0.303	0.550
Growth habit	0.991	1.004	Leaf apex shape	0.647	0.743
Young shoot color	1.429	1.240	Calyx pubescence	0.417	0.553
Young shoot pubescence	1.215	1.292	Flower diameter	2.001	2.136
Length of 'three and a bud'	2.053	2.083	Ovary pubescence	0.207	0.210
Weight of 100 'three and a bud'	1.906	1.910	Style length	1.963	2.080
Leaf attitude	1.197	1.194	Number of style splitting	0.199	0.262
Leaf size	1.233	1.265	Relative height between gynoecium and androecium	0.811	0.843
Leaf shape	0.795	0.898	Processing suitability	0.967	0.894
Leaf color	0.855	0.914	Water extracts	2.046	2.096
Leaf upper surface	1.081	1.084	Caffeine	2.061	2.153
Leaf cross section	0.784	0.732	Tea polyphenols	2.078	2.202
Leaf texture	1.045	0.999	Amino acids	2.080	2.169
Sharpness of leaf serration	0.987	0.995	Ratio of polyphenols/amino acids	1.808	1.922
Density of leaf serration	0.959	0.947	Mean±*SD*	1.184±0.602	1.224±0.613

of initial collection, indicating that some genetic redundancy was eliminated from the core collections, and the level of genetic diversity was improved.

Adequate and proper sampling that was essential in developing a representative core collection should consider the conservation of phenotypic associations arising out of coadapted gene complexes (Ortiz et al., 1998). Phenotypic correlations were calculated between the nine quantitative traits in the core collection as well as the initial collection, separately. In the initial collection, 22 correlations were significant at $P = 0.05$ whereas in the core collection, and 20 were significant at $P = 0.05$ level (Table 7). The difference between initial and core collections mainly existed in the TP/AA to LB and CAF, TP to FD, and WB to FD, while the other phenotypic correlations were consistent. This indicated a fairly good representation of coadapted gene complexes from the initial collection to the core collection.

The coincidence rate of index numbers (CRIN%) refers the proportion of the retained phenotypic index in the core collection to the total phenotypic index in the initial collection. The higher the proportion of the index number reserved, the stronger the representation of core collection indicating no loss of traits (Zhao et al., 2005). The analysis of the coincidence rate of index numbers except WB (90%) indicated that the core collection retained sufficient variations of initial collection (CRIN% = 99.7%) (Table 8).

Discussion

One of the major challenges today for genebank curators is to ensure effective and economical ways of conserving diversity. Duplication redundancy in collections has always been a major challenge faced by the curators (Mahalakshmi et al., 2007). Core collections represent the entire diversity in crops with limited use to those users who are not interested in entire diversity but rather in accessions meeting a specific diversity of traits or domain. The most important criterion for a good core collection is its representation of the entire collection, which can be evaluated by two parameters, namely, to what degree does the core collection maintain the full range of genetic variation of the entire collection and how close does

Table 7 Phenotypic correlations between quantitative traits in initial (below diagonal) and core collections (above diagonal)

Trait	LB	WB	FD	SL	WE	CAF	TP	AA	TP/AA
LB	—	0.722**	0.244**	0.160*	0.271**	0.243**	0.237**	0.075	0.064
WB	0.744**	—	0.133	0.028	0.341**	0.258**	0.313**	−0.014	0.164*
FD	0.168**	0.165**	—	0.615**	−0.053	−0.059	0.028	0.002	−0.043
SL	0.088**	−0.048	0.530**	—	−0.078	−0.100	−0.077	0.097	−0.137
WE	0.334**	0.478**	0.055	−0.018	—	0.310**	0.589**	−0.097	0.291**
CAF	0.344**	0.420**	0.010	−0.012	0.480**	—	0.277**	0.305**	−0.229**
TP	0.300**	0.412**	0.136**	−0.001	0.639**	0.356**	—	−0.309**	0.554**
AA	0.057	−0.042	0.018	0.040	0.022	0.139**	−0.248**	—	−0.703**
TP/AA	0.064*	0.193**	0.013	−0.037	0.239**	−0.039	0.539**	−0.775**	—

*: significance of phenotypic correlations at $P = 0.05$. **: significance of phenotypic correlations at $P = 0.01$.

Table 8 Coincidence rate of index numbers (CRIN) of core collection

Index	CRIN (%)	Index	CRIN (%)
Country of origin	100.0	Sharpness of leaf serration	100.0
Province of origin	100.0	Density of leaf serration	100.0
Species	100.0	Depth of leaf serration	100.0
Germplasm type	100.0	Leaf base shape	100.0
Propagating type	100.0	Leaf apex shape	100.0
Plant type	100.0	Calyx pubescence	100.0
Growth habit	100.0	Flower diameter	100.0
Young shoot color	100.0	Ovary pubescence	100.0
Young shoot pubescence	100.0	Style length	100.0
Length of 'three and a bud'	100.0	Number of style splitting	100.0
Weight of 100 'three and a bud'	90.0	Relative height between gynoecium and androecium	100.0
Leaf attitude	100.0	Processing suitability	100.0
Leaf size	100.0	Water extracts	100.0
Leaf shape	100.0	Caffeine	100.0
Leaf color	100.0	Tea polyphenols	100.0
Leaf upper surface	100.0	Amino acids	100.0
Leaf cross section	100.0	Ratio of polyphenols/ amino acids	100.0
Leaf texture	100.0	Mean	99.4

its pattern of variation resemble that of the entire collection (Ortiz et al., 1998). Brown (1989) believed that less than 30% significant difference ratio in average and range of all traits between core and overall samples and 70% average rate of all the traits of the amplitude of core samples higher than the amplitude of the whole resource groups traits could come to a conclusion that the core collection basically represent the genetic diversity of the original resources. In this study, in comparison with the means, variances, frequency distribution, phenotypic correlation, Shanon-Weaver diversity index, and coincidence rate of index numbers between the core and the initial collections, we found that there were no significant differences in the mean and range for the traits studied, and frequency distributions were not affected due to proportional sampling. The core collection also retained the overall phenotypic diversity present in the initial collection. The representative of the core collections reached more than 90%, which was found to be representative of the initial collection of tea plant. Core collection is a good candidate for the evaluation of new traits because the core collection can improve the efficiency of identifying valuable genes in a germplasm collection (Holbrook et al., 2000). This core collection reduces the number of entries to be evaluated and provides a working collection of tea plant germplasm that can be used in extensive examination for all economically important traits. The multidisciplinary evaluation of this core collection will contribute to identifying useful parents for improvement programs so as to enhance the use of genetic resources for improving quantitative traits. In addition, this core collection can be used for molecular characterization research, since the extent of diversity can be inferred for the initial collection.

Core collections are not static but dynamic to accommodate the various needs (Jaradat, 1995). As the origin center of the tea plant, China has the most abundant tea genetic resources in the world (Yu, 1986). Therefore, with the enriching of germplasms and the changing of genetic research and goals of breeding needs in the future, there exists dynamic exchange and adjustment between the core and the retained collection, rather than unchangeable. With the continual enrichment and abundance of identification and evaluation data of tea germplasms as well as the emergence of new breeding materials and innovation of germplasms, coupled with the continuous adjustment of breeding objectives, the core collection should be timely adjusted and improved in order to remain dynamic and practical in meeting the needs of breeders as well as farmers, such as the pest and disease resistant germplasms, cold resistance germplasms, more efficient use of fertilizer germplasms, etc. At the same time, as the established core collection also contains some of the germplasms not yet identified systematically, therefore, this part of the genetic resources should be identified and evaluated in order to improve the core collection database. Currently, we are using molecular markers to analyze the genetic diversity and genetic structure of the established core collection. This would remove the redundancy of resources and further reduce the group size to establish mini core collection as well as make the core collection more practical (Zhao et al., 2008).

Acknowledgements

This study was supported by the earmarked fund for Modern Agro-industry Technology Research System (Tea), China (CARS-23),

Zhejiang Natural Science Foundation Projects, China (No. Y305124), and the National Science and Technology Foundation Platform Project, China (No. 2005DKA21002-08). The authors thank Dr. Sule S. Omolaja at the Cocoa Research Institute of Nigeria and Dr. Lei-Ying Zheng at the Institute of Botany of the Chinese Academy of Sciences, for their critical reviews and constructive comments.

References

Balfourier F, Roussel V, Strelchenko P, Exbrayat-Vinson F, Sourdille P, Boutet G, Koenig J, Ravel C, Mitrofanova O, Beckert M, Charmet G (2007). A worldwide bread wheat core collection arrayed in a 384-well plate. Theor Appl Genet, 114(7): 1265–1275

Brown A H D (1989). Core collections: a practical approach to genetic resources management. Genome, 31(2): 818–824

Chen L, Yang Y J, Yu F L (2004). Tea germplasm research in China: recent progresses and prospects. Journal of Plant Genetic Resources, 5(4): 389–392 (in Chinese)

Chen L, Yang Y J, Yu F L (2005). Descriptors and Data Standard for Tea (*Camellia* spp.). Beijing: Chinese Agricultural Press (in Chinese)

Chen L, Yu F L, Tong Q Q (2000). Discussions on phylogenetic classification and evolution of Sect. *Thea*. Journal of Tea Science, 20 (2): 89–94 (in Chinese)

Frankel O H (1984). Genetic perspective of germplasm conservation. In: Arber W, Llimensee K, Peacock W J, Starlinger P, eds. Genetic Manipulations: Impact on Man and Society. Cambridge: Cambridge University Press, 161–170

Holbrook C C, Timper P, Xue H Q (2000). Evaluation of the core collection approach for identifying resistance to *Meloidogyne arenaria* in peanut. Crop Sci, 40(4): 1172–1175

Hu J, Zhu J, Xu H M (2000). Methods of constructing core collections by stepwise clustering with three sampling strategies based on the genotypic values of crops. Theor Appl Genet, 101(1–2): 264–268

Jaradat A A (1995). The dynamics of a core collection. In: Hodgkin T, Brown A H D, van Hintum T J L, Morales E A V, eds. Core Collection of Plant Genetic Resources. IPGRI, A Wiley-Sayce Publication, 179–186

Keuls M (1952). The use of the 'Studentized range' in connection with an analysis of variance. Euphytica, 1(2): 112–122

Levene H (1960). Robust tests for equality of variances. In: Olkin I, ed. Contributions to Probability and Statistics: Essays in Honour of Harold Hotelling. Stanford: Stanford University Press, 278–292.

Li C T, Shi C H, Wu J G, Xu H M, Zhang H Z, Ren Y L (2004) . Methods of developing core collections based on the predicted genotypic value of rice (*Oryza sativa* L.). Theor Appl Genet, 108(6): 1172–1176

Li J, Jiang C J (2004). Preliminary construction of core germplasm of *Camellia sinensis* in China. Journal of Anhui Agricultural University, 31(3): 282–287 (in Chinese)

Li Y, Shi Y, Cao Y, Wang T (2004). Establishment of a core collection for maize germplasm preserved in Chinese National Genebank using geographic distribution and characterization data. Genet Resour Crop Evol, 51(8): 845–852

Mahalakshmi V, Ng Q, Lawson M, Ortiz R (2007). Cowpea (*Vigna unguiculata* L. Walp.) core collection defined by geographical, agronomical and botanical descriptors. Plant Genet Resour: Characterization Util, 5(3): 113–119

Newman D (1939). The distribution of range in samples from a normal population expressed in terms an independent estimate of standard deviation. Biometrika, 31(1–2): 20–30

Ortiz R, Ruiz-Tapia E N, Mujica-Sanchez A (1998). Sampling strategy for a core collection of *Peruvian quinoa* germplasm. Theor Appl Genet, 96(3–4): 475–483

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, Liu L H (2003). Establishment of Chinese soybean (*G. max*) core collection I. sampling strategy. Scientia Agricultura Sinica, 36(12): 1442–1449 (in Chinese)

SAS Institute (1989). SAS User's Guide: Statistics. State University Press, Cary, NC, USA

Shannon C E, Weaver W (1949). The Mathematical Theory of Communication. Urbana: University of Illinois Press

Tea Research Institute Chinese Academy of Agricultural Sciences (1986). Planting Science of Chinese Tea Plant. Shanghai: Shanghai Scientific and Technical Publishers, 36–49 (in Chinese)

Wang X C, Liu Z, Yao M Z, Ma C L, Chen L, Yang Y J (2009). Sampling strategy to establish a primary core collection of Chinese tea germplasms. Journal of Tea Science, 29(2): 157–165 (in Chinese)

Wang X C, Yao M Z, Ma C L, Chen L (2008). Analysis and evaluation of biochemical components in bitter tea plant germplasms. Chinese Agricultural Science Bulletin, 24(6): 65–69 (in Chinese)

Ward Jr J H (1963). Hierarchical grouping to optimize an objective function. J Am Stat Assoc, 58(301): 236–244

Yu F L (1986). Discussion on the originating place and the originating center of tea plants. Journal of Tea Science, 6(1): 1–8 (in Chinese)

Zhao L, Dong Y, Liu B, Hao S, Wang K, Li X (2005). Establishment of a core collection for the Chinese annual wild soybean (*Glycine soja*). Chin Sci Bull, 50(10): 989–996

Zhao L P, Liu Z, Chen L, Yao M Z, Wang X C (2008). Generation and characterization of 24 novel EST derived microsatellites from tea plant (*Camellia sinensis*) and cross-species amplification in its closely related species and varieties. Conserv Genet, 9(5): 1327–1331