Chinese Science Bulletin

Genetic diversity and construction of core collection in Chinese wheat genetic resources

HAO ChenYang, DONG YuChen, WANG LanFen, YOU GuangXia, ZHANG HongNa, GE HongMei, JIA JiZeng & ZHANG XueYong†

Key Laboratory of Crop Germplasm and Biotechnology, Ministry of Agriculture/The National Key Facility for Crop Gene Resources and Genetic Improvement/Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing 100081, China

Genetic diversity among 5029 accessions representing a proposed Chinese wheat core collection was analyzed using 78 pairs of fluorescent microsatellite (SSR) primers mapped to 21 chromosomes. A stepwise hierarchical sampling strategy with priority based on 4×10^5 SSR data-points was used to construct a core collection from the 23090 initial collections. The core collection consisted of 1160 accessions, including 762 landraces, 348 modern varieties and 50 introduced varieties. The core accounts for 23.1% of the 5029 candidate core accessions and 5% of the 23090 initial collections, but retains 94.9% of alleles from the candidate collections and captures 91.5% of the genetic variation in the initial collections. These data indicate that it is possible to maintain genetic diversity in a core collection while retaining fewer accessions than the accepted standard, i.e., 10% of the initial collections captured more than 70% of their genetic diversity. Estimated genetic representation of the core constructed by preferred sampling (91.5%) is much higher than that by random sampling (79.8%). Both mean genetic richness and genetic diversity indices of the landraces were higher than those of the modern varieties in the core. Structure and principal coordinate analysis revealed that the landraces and the modern varieties were two relatively independent subpopulations. Strong genetic differentiation associated with ecological environments has occurred in the landraces, but was relatively weak in the modern cultivars. In addition, a mini-core collection was constructed, which consisted of 231 accessions with an estimated 70% representation of the genetic variation from the initial collections. The mini-core has been distributed to various research and breeding institutes for detailed phenotyping and breeding of genetic introgression lines.

Triticum aestivum L., SSR, genetic diversity, core collection

China is one of the diversity centers of common wheat (*Triticum aestivum* L.)[1]. The long cultivation history and artificial selection in different ecological zones have created abundant genetic variation, which was collected and preserved in the Chinese National Gene Bank. About 2×10^4 domesticated wheat accessions, approximately half landraces and half modern varieties, are conserved in the national gene bank. These resources provide basic materials for genetic research and breeding. However, the huge number of accessions also exerts great difficulties in conservation, evaluation, identification and utilization[2].

In the 1980s, Frankel[3,4] and Brown[5,6] proposed and developed the concept of core collection. Many countries and international agricultural research centers have given this concept strong attention and core collections have been constructed in several crops since that time[7−12]. Using the fewest accessions to retain the greatest genetic diversity of the initial collections is the key point in construction of a core collection. Therefore,

Received December 24, 2007; accepted March 11, 2008
doi: 10.1007/s11434-008-0212-x
†Corresponding author (email: xueyongz@public.bta.net.cn)
Supported by the National Basic Research Program of China (Grant Nos. G19980202 and 2004CB117202)

reasonable selection of appropriate sampling scale is an important and disputed problem[5,13]. Previous research showed that the proper sampling scale ranges were 5 to 30% in different crops, resulting in retention or more than 70% of the genetic diversity from the initial pools[14−17]. Recently, Balfourier et al.[18] used 372 core entries to represent 98.2% of the genetic diversity from 3942 wheat germplasm resources from 73 countries.

The core collection of Chinese common wheat was initiated in 1999. We have constructed a candidate core collection comprising 5029 common wheat accessions by clustering and square root sampling strategies[1,19] from 23090 initial collections based on phenotyping 21 agronomic and botanic traits. Here, we discuss sampling strategy and construction of the core collection based on SSR data from the candidate core collections. The core collection will provide a platform for verification of agronomically important traits, diversity analysis of functional genes, and large-scale association mapping[20]. Geographic distribution of genetic diversity was also evaluated in the core collection.

1 Materials and methods

1.1 Plant materials

The Chinese wheat candidate core collection, composed of 5029 accessions, including 3373 landraces and 1656 modern varieties selected from 23090 initial collections, was used in this research. The candidate retains 96.4% of the genetic diversity from the initial collections, estimated from the passport data in the gene bank (Table 1)[1]. Additionally, 70 introduced varieties were included to analyze genetic structure and their influence on national wheat breeding because they have been cultivated

Table 1 Sampling percentages and allelic representation of Chinese wheat core collection entries from the ten ecological regions[a)]

Wheat regions	Type	Candidate CC	CC	Sampling scale (%)	No. of alleles in candidate CC	No. of alleles in CC	Alleles retained (%)	Variation representation (%)
Northern winter Wheat region I	landraces	388	79	20.4	998	804	80.6	77.7
	modern varieties	227	56	24.7	868	691	79.6	76.7
Yellow and Huai River valley	landraces	797	207	26.0	1088	952	87.5	84.4
Winter wheat region II	modern varieties	481	117	24.3	993	801	80.7	77.8
Low and middle Yangtze River valley	landraces	693	105	15.2	1000	757	75.7	73.0
Winter wheat region III	modern varieties	295	51	17.3	831	631	75.9	73.2
Southwestern winter Wheat region IV	landraces	521	128	24.6	1068	912	85.4	82.3
	modern varieties	183	40	21.9	744	560	75.3	72.6
Southern winter Wheat region V	landraces	173	30	17.3	793	515	64.9	62.6
	modern varieties	43	8	18.6	574	325	56.6	54.6
Northeastern spring Wheat region VI	landraces	84	22	26.2	770	556	72.2	69.6
	modern varieties	148	22	14.9	716	453	63.3	61.0
Northern spring Wheat region VII	landraces	171	45	26.3	878	677	77.1	74.3
	modern varieties	53	9	17.0	632	370	58.5	56.4
Northwestern spring wheat region VIII	landraces	229	70	30.6	936	799	85.4	82.3
	modern varieties	110	32	29.1	764	608	79.6	76.7
Qinghai-Tibet spring-Winter wheat region IX	landraces	151	46	30.5	947	757	79.9	77.0
	modern varieties	13	4	30.8	403	189	46.9	45.2
Xinjiang winter-Spring wheat region X	landraces	166	30	18.1	981	682	69.5	67.0
	modern varieties	33	9	27.3	569	379	66.6	64.2
Chinese varieties	landraces	3373	762	22.6	1531	1408	92.0	88.7
	modern varieties	1586	348	21.9	1336	1158	86.7	83.6
Introduced varieties	modern varieties	70	50	71.4	798	737	92.4	89.1
Total		5029	1160	23.1	1641	1557	94.9	91.5

a) Wheat ecological regions and distribution of accessions are according to Zhuang[21] and Chinese wheat genetic resource catalogs[22]. CC, core collection.

in large scale or used as founder genotypes in breeding.

1.2 Microsatellite data collection and analysis

Seventy-eight microsatellite (SSR) loci, screened for high levels of polymorphism and mapped to a location on the wheat chromosomes, were used to genotype the candidate core collection. An ABI 3700 DNA Analyzer was used to detect the amplification products (alleles) of the fluorescence-labeled SSR primers[10,23−26]. The raw data were processed with Genescan3.7, Genotyper3.7 (http://www.appliedbiosystems.com) and Excel software.

1.3 Data collection and statistics

The length and presence of the alleles were scored individually, and scores were recorded as 1 for present and 0 for absent. Genetic diversity of each locus (i) was expressed by allelic richness ($\sum A_{ij}$), and polymorphic information content (PIC), $\mathrm{PIC}_i=1-\sum_{j=1}^{l} p_{ij}^2$, where j represents each allele at the i locus, p_{ij} is the frequency of each allele at i locus, l means the total alleles at i locus. Genetic diversity was computed by $H_t=\sum_{i=1}^{k}\mathrm{PIC}_i/n$, where n is the total number of loci studied[27].

To determine the minimum sampling rate necessary to represent 70% of the diversity of the initial collection, random sampling simulation at each sampling level was carried out by computer simulations. Core entries were randomly sampled ten times from the candidate core collection. The mean allele number retained by the ten simulations was regarded as the capacity for the responding sampling rate. This simulation was carried out in the landraces and the modern varieties separately.

1.4 Strategy for construction of core collection and decision of the entries

The sampling rate of the core collection was set to 5% of the initial collection (1160 accessions), or 23.1% of the candidate core collection. The number of entries sampled from each ecological region was adjusted according to the region's genetic diversity. Entry number was raised for those regions with higher genetic diversity and reduced in ones with lower genetic diversity. All the candidate core accessions were divided into two major populations, the landraces and the modern varieties. Then each of the two populations was divided into ten groups based on their ecological regions[21]. Cluster analysis for each group was carried out based on the SSR data. Entries for the core collection were selected from each major branch of the cluster trees. Very well known varieties or accessions with unique SSR alleles were given priority in choice of entries[11,28]. This sampling strategy is called stepwise hierarchical sampling strategy with priority. To verify the advantage of this sampling method, we compared it with random sampling strategy and partial random sampling strategy using the same number of entries (i.e. 1160 accessions). For the random sampling strategy, core entries were selected randomly. For the partial random sampling strategy, sampling number from each ecological region was adjusted according to its genetic diversity, but the entries were randomly selected.

The genetic similarities between accessions were calculated using the DICE coefficient (NTSYS-pc version 2.1)[29]. Cluster analysis was performed based on the similarity matrices with the Un-weighted Pair Group Method with Arithmetic dendrogram (UPGMA).

1.5 Representation calculation and genetic structure analysis of the core collections

Since the candidate core collection retains 96.4% of the morphological and agronomic variation from the initial collection[1], genetic representation of the core collection was calculated as 96.4% × percentage of alleles kept[10].

Genetic structure of the core collection was analyzed by Structure2.1 software[30]. Principal coordinate analysis (PCO) in the NTSYS-pc version 2.1[29] was also used for structure analysis in the core collection.

2 Results and analysis

2.1 Sampling strategy and genetic representation of core collection

(i) Alleles of each region at the same sampling number. After genotyping the 5029 candidate entries at 78 SSR loci, a total of 1641 alleles were amplified. The landraces and the modern varieties possessed 1531 and 1336 alleles, respectively. Large differences existed for allele number among regions and between the landraces and modern varieties. To minimize the influence of accession number from different ecological regions on allelic number, 78 landraces and 42 modern varieties from the candidate core collection were selected randomly from each region for comparing allelic richness

among different regions. Because the candidate entries of the modern varieties from Qinghai-Tibet spring-winter wheat and Xinjiang winter-spring wheat regions were fewer than 42, all of the candidates were used in this comparison (Figure 1). The landraces have the highest allele numbers in Qinghai-Tibet spring-winter wheat region (IX) and Xinjiang winter-spring wheat region (X), and have the fewest alleles in Low and middle Yangtze River valley winter wheat region (III) and Southern winter wheat region (V). For the modern varieties, Yellow and Huai River valley winter wheat region (II) and Northwestern spring wheat region (VIII) have the richest alleles, and region IX has the least.

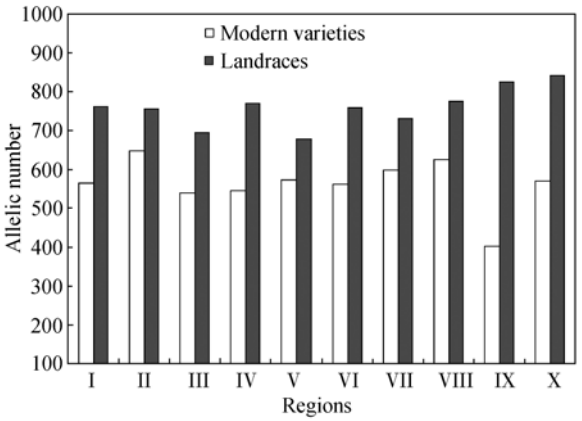

Figure 1 Allelic number of landraces and modern varieties respectively in the ten regions. I, Northern winter wheat region; II, Yellow and Huai River valley winter wheat region; III, Low and middle Yangtze River valley winter wheat region; IV, Southwestern winter wheat region; V, Southern winter wheat region; VI, Northeastern spring wheat region; VII, Northern spring wheat region; VIII, Northwestern spring wheat region; IX, Qinghai-Tibet spring-winter wheat region; X, Xinjiang winter-spring wheat region.

(ii) Sampling rate and genetic representation. To estimate the minimum number of collections needed to retain 70% of the genetic diversity from the initial collections, random sampling simulation was carried out in both the landrace and modern variety populations. We found that the relationship between allelic representation and sampling number of collections was not linear in either the landraces or modern varieties (Figure 2). To capture more than 70% of the genetic diversity from the landraces, the number of the accessions should be more than 350, accounting for 10.4% of the candidate core collection and 3.2% of the initial accession. In the modern varieties, to reach the same goal, the number of the accessions should be more than 300, which is 18.1% of the candidate core collection and 2.7% of the initial collection (Figure 2).

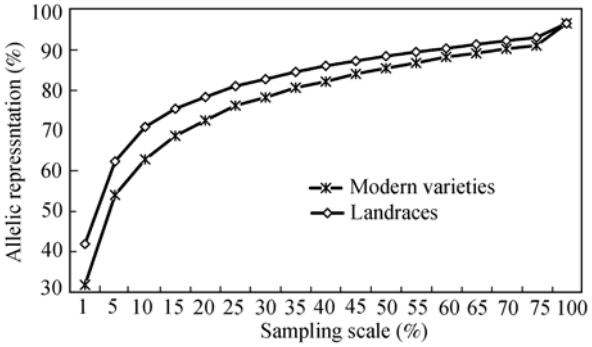

Figure 2 Mean percentage of alleles retained at various randomly sampling scales in the candidate core collections.

Considering the obvious genetic diversity differences among geographic and ecological regions (Figure 1), the sampling number was adjusted for each region. Raising the sampling number from the region with higher genetic diversity would keep the genetic representation higher than 70% while using fewer entries from the whole country. Considering common alleles among different ecological regions and between the landraces and the modern varieties, a larger sampling scale (5% of the initial collections, and 23.1% of candidate core entries) was adapted in the core collection construction to maintain about 70% of the genetic diversity from each region. Using the stepwise hierarchical sampling strategy with priority, the core entries were determined in each region based on the cluster results. Entries from each of the ten regions formed the national core collection.

The core collection consists of 1160 accessions, accounting for 23.1% of the original candidate core collection and 5% of the initial collection, respectively, and keeps 94.9% of the alleles detected in the candidate core collections. It should retain 91.5% of the genetic diversity of the initial collection (Table 1), which is considerably higher than the internationally accepted standard level, i.e., using 10% of the initial collection to represent more than 70% of the genetic diversity[3—5].

(iii) Genetic representations of preferred sampling vs random sampling. To verify that preferred sampling is better than complete random sampling or partial random sampling in constructing the core collection, we compared the genetic representation of the 1160 entries selected by these three sampling methods. The data in Figure 3 clearly showed that the core collection selected by the preferred sampling method represented 91.5% of

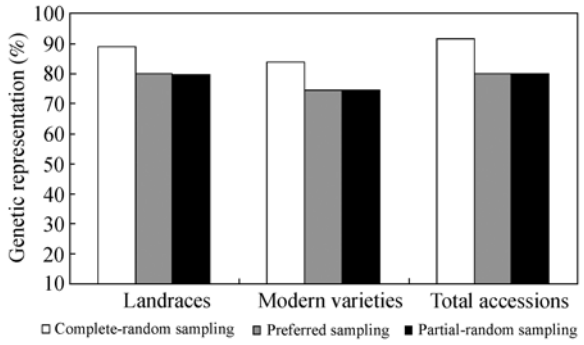

Figure 3 Genetic representation difference of the core collection set up by preferred sampling, random sampling and partial random sampling.

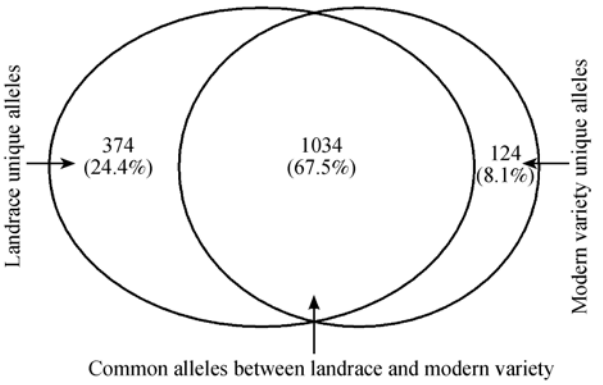

Figure 4 Distribution of alleles in landraces and modern varieties in Chinese wheat core collections.

the genetic diversity, while the entries selected by complete random sampling or partial random sampling represented much lower genetic diversity (79.8%).

2.2 Genetic diversity of the core collection

(i) Diversity of landraces vs modern varieties. In total, 1557 alleles were amplified at the 78 SSR loci in the 1160 core collection. The number of alleles of landraces and modern varieties were 1408 and 1158 respectively. Within the 1557 alleles, 374 (24.4% of the global) and 124 (8.1% of the global) alleles were only detected in the landraces and the modern varieties, respectively, and 1034 alleles (67.5% of the global) were shared by the landraces and the modern varieties (Figure 4). These data demonstrate that the landraces have more unique alleles. In other words, the landraces have higher genetic diversity than the modern varieties.

The 78 pairs of SSR primers produced 737 alleles in the 50 introduced varieties (Table 1). Of these 737 alleles, 682, 699 and 669 alleles were also detected in the Chinese modern varieties, the Chinese landraces, and in both, respectively. Only 25 unique alleles were found in the introduced varieties.

(ii) Genetic structure of Chinese wheat core collections. Genetic structure of the 762 landraces and 348 modern varieties in the core was analyzed using Structure2.1[30]. The accessions were clustered into two basic groups (landraces and modern varieties) basically at K = 2 (Figure 5). Principal coordinate analysis (PCO) of the core further revealed that Chinese landraces and modern varieties were two relatively independent sub-populations (Figure 6)[10,25,26].

Separate structure analysis of the landraces and modern varieties in the core revealed that strong genetic differentiation occurred in the landraces, which was associated with long-term ecological selection. However, genetic differentiation was relatively weak in the modern cultivars (Figures 7 and 8). Only varieties from regions V & X showed differentiation compared with those from the other eight regions, which was related with their underdeveloped breeding levels. Widespread use of founder genotypes across the country has weakened the genetic borders among ecological regions.

(iii) Genetic diversity of core collections in each ecological region. There were large differences in allelic number among the Chinese ten ecological regions. Allelic number ranged from 515 to 952 for the landraces, and 189 to 801 for the modern cultivars, further demonstrating the higher number of alleles in the landraces compared to the modern varieties.

Genetic diversity of the core collection selected from each ecological region was assessed with average ge-

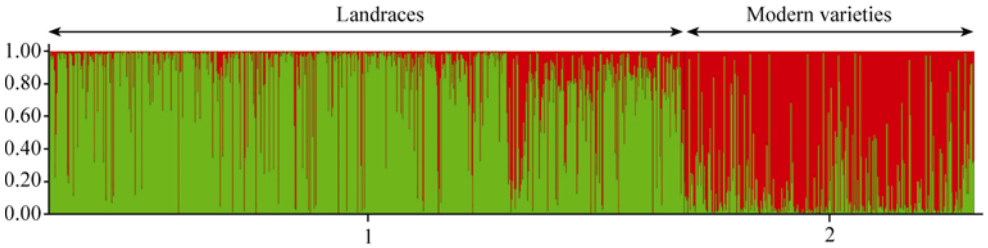

Figure 5 Genetic structure in Chinese wheat core collections at K = 2.

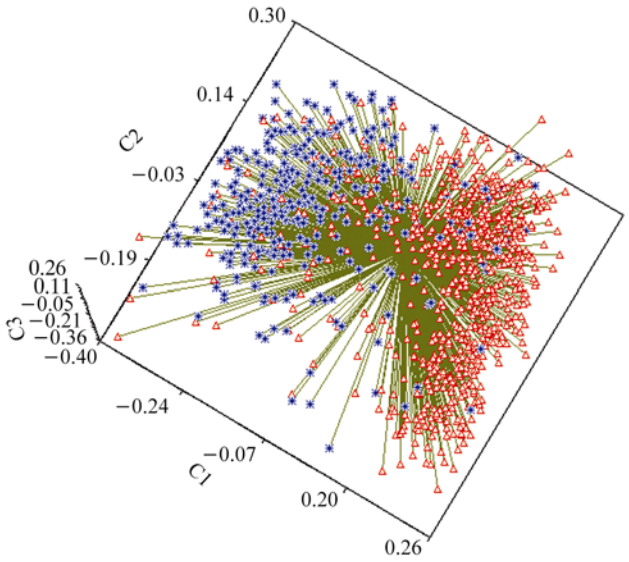

Figure 6 PCO of Chinese wheat core collections. *, Modern varieties; △, landraces.

netic richness and genetic diversity indices (Figure 9). For the landraces, region II had the highest genetic richness, followed by region IV. Regions I and VIII had relatively higher genetic richness, while regions V and VI had the lowest. These results further confirmed that regions II and IV were the diversity centers of Chinese common wheat landraces.

Region II possessed the highest genetic diversity in the modern varieties. The modern varieties from region I had the second highest genetic diversity. The lowest genetic diversities of the modern varieties were found in regions IX and V. This result reflected that wheat breeding in regions I and II were stronger, but relatively weaker in regions IX and V. Analysis of the correlations for genetic diversity between the core collection and the candidate core collection from different ecological regions showed that there was high correlation for both mean genetic richness (r = 0.976, P<0.01) and genetic diversity indices (r = 0.942, P<0.01).

It should be emphasized that it was necessary to analyze both mean genetic richness and genetic diversity indices (H_t) in the evaluation of genetic diversity because H_t cannot fully reflect genetic diversity richness, especially when the number of accessions is not high. For example, the modern varieties from region IX had a very low value of mean genetic richness, but quite high H_t. In fact, genetic diversity of the modern varieties in this region was low (Figures 9 and 10).

3 Discussion

3.1 Grouping in core collection construction

Previous research indicated that the sampling strategy based on stepwise hierarchical sampling was helpful to retain the maximum genetic diversity with the least repetition in the core collection[31]. In construction of the Chinese wheat core collection, two-stepwise grouping methods were used. All accessions were divided into two populations (landraces and modern varieties). Then the two populations were divided into ten groups ac-

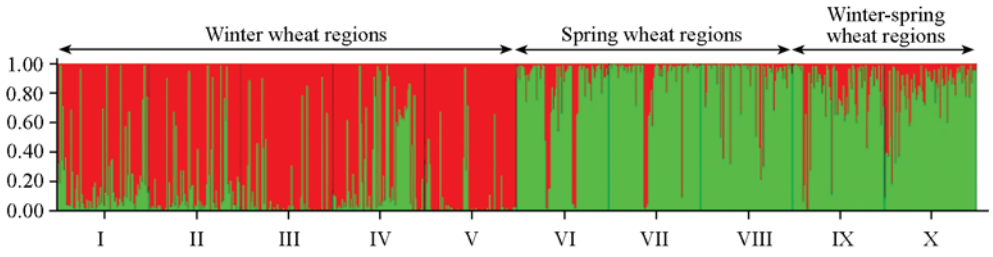

Figure 7 Population structure of landraces in the core from the 10 ecological regions at K = 2.

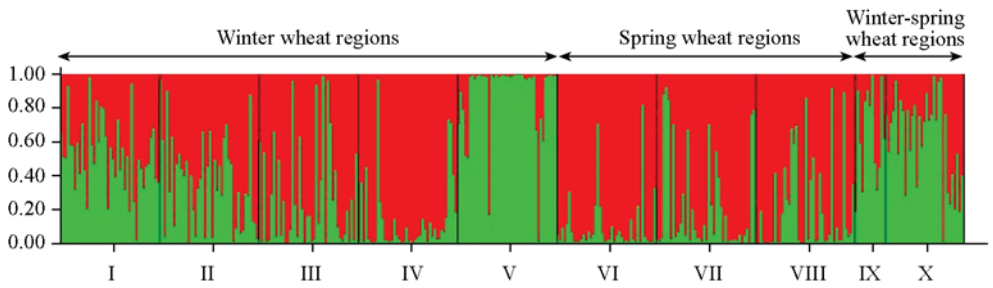

Figure 8 Population structure of modern varieties in the core from the 10 ecological regions at K = 2.

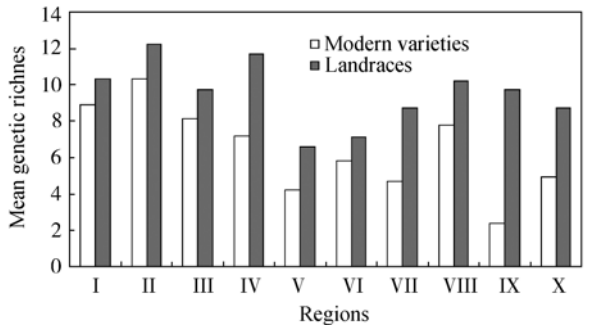

Figure 9 Mean genetic richness in the ten ecological regions evaluated in the core.

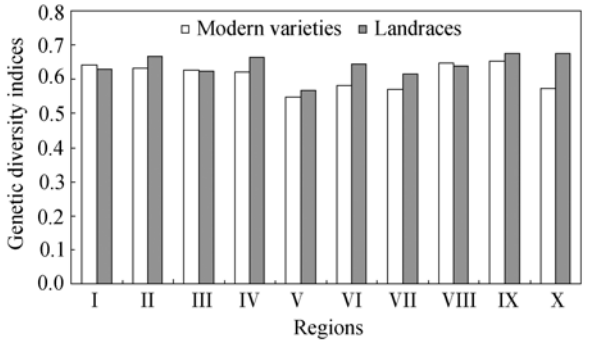

Figure 10 Genetic diversity indices in the ten ecological regions evaluated in the core.

cording to their ecological regions[21]. This grouping method was consistent with the Chinese wheat cultivation practice. Genetic structure analysis of Chinese wheat showed that the landraces and the modern varieties were two relative independent genetic populations (Figures 5 and 6).

China has more than 4 millennia history of wheat cultivation, and most landraces formed isolates because of transportation limitation in the earlier time. Scientific breeding in China can be traced back to 50—90 years[21]. The history of Chinese wheat breeding shows that new varieties were selected from landraces in the early period; a small number of introduced varieties were popularized directly in the 1920s. Varieties derived from crosses between landraces and introduced varieties were used in wheat production from the 1930s to 1940s. Varieties released in the 1960s were mainly derived from crosses (or multi-crosses) between the introduced varieties and the Chinese landraces or their selections. Since the 1970s, most varieties were derived from crosses between Chinese modern varieties[21]. Therefore, it was understandable that the landraces and the modern varieties were two sub-populations in genetic structure. Thus, in the Chinese wheat core collection construction, it was reasonable to classify all accessions into two groups, the landraces and the modern varieties, in the first step. Because Chinese wheat production was divided into ten ecological regions[21], it was suitable to further group the landraces and the modern varieties according to their ecological classification in the second step.

3.2 Sampling strategy for construction of core collection

In general, the number of core collections could take account of 5% to 10% of the basic collection, and genetic representation should be higher than 70%[5,6]. Considering the large difference in genetic diversities among the ten ecological regions (Figures 1, 9 and 10), sampling number was raised for those regions with higher genetic diversity and fewer initial collections, and was reduced for those regions with lower genetic diversity and large number of initial collections. Priority was given to those accessions either famous or important in breeding, production, or that conveyed unique alleles. Through preferred sampling of the entries, the core collection, accounting for 5% of the initial collection, could represent 91.5% of the genetic diversity of the initial collection, much higher than that of randomly sampled entries (79.8%). Though it was very successful in Chinese wheat core collection construction, the adjustment of sampling number was very difficult in practical operation. For example, genetic representation was only 70% even using as high as 26.2% of the landraces in the Northeastern spring wheat region. Because there are a small number of modern varieties in the Qinghai- Tibet spring-winter wheat region, even when the sampling percentage was raised to 30.8%, the representation was only 45.2%. In these regions the sampling rate could not be reduced too much. However, if sampling number of the landraces was slightly reduced in the Northwestern spring wheat region and the Northern spring wheat region and was raised in the Xinjiang winter-spring wheat region, and sampling number of the modern varieties was raised in the Northern spring wheat region and the Northeastern spring wheat region, the genetic representation would be higher.

3.3 Significance of core collection

(i) A good platform for marker/trait association and genetic analysis. In construction of the Chinese wheat core collection, 400000 SSR data-points and 100000 phenotypic data-points were collected. These data-points

provide a good platform for marker/trait association analysis for mapping important agronomic traits and revealing the evolution of wheat cultivars in China[19,20,32]. In addition, marker/trait association analysis has already detected a number of important genomic regions, which will provide targets for fine mapping and haplotyping analysis in segregating populations or genetic introgression lines.

Based on the SSR data obtained from the candidate core collection, our analysis showed that variety diversity has declined very quickly since the 1950s in China[32]. Ecological distribution of genetic diversity of Chinese wheat germplasm resources was analyzed in the present study. This is very important for determining genetic diversity centers of common wheat in China (Figures 9 and 10). For the landraces, it is quite obvious that ecological region II and region IV are two centers of diversity. The role of region VIII could not be neglected in the introduction of wheat into China (Figure 9).

Among the 1557 alleles in the core, 669 alleles were shared by the landrace, the modern variety and the introduced variety entries, and 1034 alleles were shared by the landrace and the modern variety entries. These common alleles, especially those with extremely high frequency may be associated with important agronomic traits[20]. Three sub-populations have their own unique alleles, but the landrace has the most (24.4%). This indicated that there may be opportunity for discovering new alleles in the landraces.

(ii) Core- and mini-core collection providing material bases for gene discovering. After all core collections were assembled, 93.5% of all SSR alleles in the candidate core were retained. Overall, 5% of the entries (1160 accessions) captured 91.5% of the variation present in the 23090 initial accessions. Besides the core, we have also developed a mini-core collection consisting of 231 entries (1% of the initial collection), which represents 70.1% of the SSR alleles. The mini-core facilitates the genome scanning of this crop to reveal the evolutionary mechanisms of functional and repetitive DNA sequences. The mini-core also permits the rapid identification of potentially useful alleles through breeding and analyzing genomic introgression lines assisted by haplotype analysis[33]. Mini-core collections have been provided to scientists in different research fields for verification of characteristics, or to breeders for pre-breeding. Secondly, mini-core collections have been used as the donors in construction of introgression lines for gene discovery. Introgression lines with new genes are being provided to breeders for MAS breeding[34,35].

We gratefully appreciate Professors Zhuang Q S, Li Z S and Zhang Q F for their valuable comments and Dr. Joe Robin and Dr. Steven Xu of USDA- ARS for their help on English editing.

1 Dong Y S, Cao Y S, Zhang X Y, et al. Establishment of candidate core collections in Chinese common wheat germplasm. J Plant Genet Resour (in Chinese with English abstract), 2003, 4: 1－8
2 Tanksley S D, McCouch S R. Seed bank and molecular maps: Unlocking genetic potential from the wild. Science, 1997, 277: 1063－1066
3 Frankel O H. Genetic perspectives of germplasm conservation. In: Arber W, Limensee K, Peacock W J, et al, eds. Genetic Manipulation: Impact on Man and Society. Cambridge: Cambridge University Press, 1984. 161－170
4 Frankel O H, Brown A H D. Plant genetic resources today: A critical appraisal. In: Holden J H W, Williams J T, eds. Crop Genetic Resources: Conservation & Evalution. London: George Allen & Urwin Ltd, 1984. 249－257
5 Brown A H D. Core collections: A practical approach to genetic resources management. Genome, 1989, 31: 818－824
6 Brown A H D. The case for core collections. In: Brown A H D, Frankel O H, Marshall D R, et al, eds. The Use of Plant Genetic Resources. Cambridge: Cambridge University Press, 1989. 136－156
7 Hamon S, van Stolen D H. Characterization and evaluation of okra. In: Brown A H D, Frankel O H, Marshall D R, et al, eds. The Use of Plant Genetic Resources. Cambridge: Cambridge University Press, 1989. 173－196
8 Corley H C, William F A. Evaluation of a core collection to identify resistance to late leaf spot in peanut. Crop Sci, 1995, 35: 1700－1702
9 Upadhyaya H D, Bramel P J, Singh S. Development of a chickpea core subset using geographic distribution and quantitative traits. Crop Sci, 2001, 41: 206－210
10 Hao C Y, Zhang X Y, Wang L F, et al. Genetic diversity and core collection evaluations in common wheat germplasm from the Northwestern Spring Wheat Region in China. Mol Breed, 2006, 17: 69－77
11 Hu J, Zhu J, Xu H M. Methods of constructing core collections by stepwise cluster with three sampling strategies based on genotypic values of crops. Theor Appl Genet, 2000, 101 (1-2): 264－268
12 Wang J C, Hu J, Liu N N, et al. Investigation of combining plant genotypic values and molecular marker information for constructing core subsets. J Integr Plant Biol, 2006, 48: 1371－1378
13 Yonezawa K, Nomura T, Morishima H. Sampling strategies for use in stratified germplasm collections. In: Hodekin T, Brown A H D, van

Hintum T J L, et al, eds. Core Collections of Plant Genetic Resources. Chichester: Wiley-Sayce Publication, 1995. 35－54
14 Brown A H D, Grace J P, Speer S S. Designation of a "core" collection of perennial Glycine. Soybean Genetics Newsletter, 1987, 14: 59－70
15 van Hintum T J L. Comparison of marker systems and construction of a core collection in a pedigree of European spring barley. Theor Appl Genet, 1994, 89: 991－997
16 Li Z C, Zhang H L, Zeng Y W, et al. Study on sampling schemes of core collection of local varieties of rice in Yunnan, China. Sci Agric Sin (in Chinese), 2000, 33(5): 1－7
17 Ellegren H. Microsatellites: Simple sequences with complex evolution. Nat Rev Genet, 2004, 5: 435－445
18 Balfourier F, Roussel V, Strelchenko P, et al. A worldwide bread wheat core collection arrayed in a 384-well plate. Theor Appl Genet, 2007, 114: 1265－1275
19 Zhang X Y, Pang B S, You G X, et al. Allelic variation and genetic diversity at *Glu-1* loci in Chinese wheat (*Triticum aestivum* L.) germplasms. Sci Agric Sin (in Chinese), 2002, 35 (11): 1302－1310
20 Zhang X Y, Tong Y P, You G X, et al. Hitchhiking effect mapping: A new approach for discovering agronomic important genes. Sci Agric Sin (in Chinese), 2006, 39 (8): 1526－1535
21 Zhuang Q S. Chinese Wheat Improvement and Pedigree Analysis (in Chinese). Beijing: Agricultural Press, 2003
22 Institute of Crop Germplasm Resources, CAAS. Chinese Wheat Genetic Resource Catalogs. Beijing: China Agriculture Press, 1995.
23 Hao C Y, Wang L F, Dong Y C, et al. Comparison of fluorescence and silver-staining detection systems of microsatellite markers. Acta Agron Sin (in Chinese), 2005, 31: 144－149
24 Röder M S, Korzun V, Wendehake K, et al. A microsatellite map of wheat. Genetics, 1998, 149: 2007－2023
25 Zhang X Y, Li C W, Wang L F, et al. An estimation of the minimum number of SSR alleles needed to reveal genetic relationships in wheat varieties. I. Information from large-scale planted varieties and corner-stone breeding parents in Chinese wheat improvement and production. Theor Appl Genet, 2002, 106: 112－117
26 You G X, Zhang X Y, Wang L F. An estimation of the minimum number of SSR loci needed to reveal genetic relationships in wheat varieties: Information from 96 random samples with maximized genetic diversity. Mol Breed, 2004, 14: 397－406
27 Nei M. Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci USA, 1973, 70: 3321－3323
28 Hu J, Xu H M, Zhu J. A method of constructing core collection reserving special germplasm materials. J Biomathemat, 2001, 16 (3): 348－352
29 Rohlf F J. NTSYS-pc: Numerical taxonomy and multivariate analysis system, Version 2.1. NY: Exeter Software, 2000
30 Pritchard J K, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics, 2000, 155: 945－959
31 Spagonletti Z P L, Qualset C O. Evaluation of five strategies for obtaining a core subset from a large genetic resource collection of durum wheat. Theor Appl Genet, 1993, 87: 295－304
32 Hao C Y, Wang L F, Zhang X Y, et al. Genetic diversity in Chinese modern wheat varieties revealed by microsatellite markers. Sci China Ser C-Life Sci, 2006, 49: 218－226
33 Peleman, J D, van der Voort R. Breeding by design. Trends Plant Sci, 2003, 8: 330－334
34 Jia J Z. Applications of theories and approaches of plant genomics. Revi China Agricult Sci Technol (in Chinese), 1999, 2: 41－45
35 He G M, Luo X J, Tian F, et al. Haplotype variation in structure and expression of a gene cluster associated with a quantitative trait locus for improved yield in rice. Genome Res, 2006, 16: 618－626