Theor Appl Genet (2002) 106:112–117
DOI 10.1007/s00122-002-1016-z

X. Y. Zhang · C. W. Li · L. F. Wang · H. M. Wang
G. X. You · Y. S. Dong

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 cornerstone breeding parents in Chinese wheat improvement and production

Received: 14 January 2002 / Accepted: 2 May 2002 / Published online: 6 August 2002

Abstract The genetic relationships of 43 wheat varieties were analyzed with SSR markers. The materials employed included 14 cornerstone breeding parents used before 1980 and another 29 other large-scale planted varieties currently in use in China. A total of 501 different alleles were amplified, including 166 alleles of the **A** genome, 174 of the **B** genome and 161 of the **D** genome. Data obtained were used to estimate genetic similarity using the DICE coefficient, and dendrograms were constructed using the UPGMA method. The dendrogram with 501 alleles successfully differentiated all the cornerstone breeding parents and the large-scale planted varieties, and the dendogram tree was basically consistent with the pedigrees of these varieties. The correlation coefficient between the genetic distance matrix of 501 alleles and that of 450 was 0.99. Correlation coefficients among random samples of alleles suggested that 350 to 400 alleles were needed to detect genetic relationships among common wheat varieties. Correlation coefficients of a genetic similarity matrix based on 580, and those of 501 and 400, random alleles were 0.96 and 0.94, respectively. However, there were marked differences between the matrix based on the 501 alleles and those based on markers located on the **A**-, **B**- or **D**-genome independently. The correlation coefficients between the genetic distance matrix of 501 alleles and alleles within **A**, **B** or **D** genomes were 0.77, 0.76 and 0.67. The estimation of genetic similarity should be based on data from all genomes rather than any one genome.

Communicated by H.C. Becker

X.Y. Zhang (✉) · C.W. Li · L.F. Wang · H.M. Wang
G.X. You · Y.S. Dong
Key Lab of Crop Germplasm and Biotechnology,
Institute of Crop Germplasm Resources,
Chinese Academy of Agriculture Sciences (CAAS), Beijing,
China 100081
e-mail: xueyongz@public.bta.net.cn
Tel.: 010-62186630, Fax: 010-62186629

Keywords Wheat · SSR · Genetic relationships · Genetic diversity

Introduction

In China, common wheat (*Triticum aestivum* L.) is the second largest crop of economic importance. Since 1949, yield and total production have increased significantly due to the release and utilization of new varieties (Jing et al.1997). A few breeding lines have played a particularly key role in the improvement of wheat varieties and are designed as cornerstone breeding parents (see Table 1). Other materials have been cultivated, largely as commercial varieties (Jing et al. 1983, 1986, 1997; Zhang et al. 2002a; Zheng 2001), and originated from adapted progenies of the cornerstone varieties. Investigation of the genetic relationships of these cornerstone parents may be useful for future wheat improvement and breeding.

Common wheat is an allohexaploid (2n = 42) and consists of three different genomes: **A**, **B** and **D**. Bread wheat is the product of hybridization between the tetraploid wheat *Triticum turgidum* L. (2n = 28, **AABB**), and the diploid grass, *Aegilops tauschii* Coss. (2n = 14, **DD**) (Kihara 1944; McFadden and Sears 1946). The highest concentration of microsatellite loci in the existing map of wheat is on the **B** genome and the lowest is on the **D** genome. This reflects the amount of polymorphism found in the different genomes of the mapping population (Pestsova et al. 2000). And each of the three genomes should have different genetic diversity.

Some sections of DNA are characterized by the presence of a variable number of tandem repeats, which allow their use as markers, termed microsatellites (SSRs). The basic unit of these SSRs is usually one to five nucleotides long, but the number of repetitions of the basic unit is quite variable between varieties. Because of this, SSRs are highly informative, locus-specific, genetic markers (Danin-Poleg et al. 2000). Microsatellites have additional advantages in that they are abundant and evenly dispersed throughout animal and plant genomes (Hamada et al.

Table 1 List of 43 varieties of common wheat included in our study and their time used in breeding and production

Varieties	New varieties created	Year starting used in breeding	Major time used in production
Chinese Spring (CK)			
Yanda 1817[a,b]	53	1948	Around 1949
Chengduguantou[a]	29	1946	1940–60
Mazamai[a]	95	1942	No detail data
Youbaomai			1972–80
Bima 4[a,b]	68	1948	1951–60
Beijng 8[a,b]	33	1962	1968–74
Xinong 6028[a]	31	1949	1951–60
Wuyimai[a,b]	25	1951	1958–65
Mentana(Nanda 2419)[a,b]	110	1942	1949–61
Orofen[a,b]	110	1966	1964–79
Funo[a,b]	165	1957	1957–80
Abbondanza[a,b]	217	1961	1961–79
Early Premium[a]	58	1950	1950–64
Nongda 139[b]			1970–83
Jimai 21[b]			1980–87
Ganmai 8[b]			1965–81
Neixiang 5[b]			1960–75
Taishan 4[b]			1973–81
Taishan 5[b]			1980
Bonong 7023[b]			1970–82
Bonong 74-22[b]			1979–92
Zhengzhou 683[b]			1969–82
Jinan 13[b]			1979–90
Bainong 3127[b]			1978–90
Xiaoyan 6[b]			1982–96
Quality[b]			1949–62
Yangmai 1[b]			1967–79
Yangmai5[b]			1986
Yangmai 158[b]			1992
Xuzhou 14[b]			1965–75
Villa Glori[b]			1953
Ardito[b]			1942
Mara	9	1957	1959–70
St2422–464	13	1965	
Wan 7107[b]			1980–92
Emai 6			1969–83
Fan 6[b]			1974–84
Akagomughi (CK)			
Jiangdongmen[a]	50	1934	
Fengchan 3[b]			1968–78
Aimengniu	13	1978	
Youzimai[a,b]	17	1958	

[a] The 14 cornerstone lines of Chinese wheat breeding prior to 1980

[b] The annual maximum acreage reached about 10 million mu (667,000 hm^2)

1982; Beckmann and Soller 1990), are co-dominant and have a high information content, and their PCR-based detection facilitates the study of genetic diversity (Plaschke et al.1995), gene mapping (Röder et al. 1998; Pestsova et al. 2000) and testing of authenticity of genetic stocks in common wheat (Pestsova et al. 2000).

Construction of a stable phenetic dendrogram is a prerequisite for genetic diversity research and establishing core collections. The minimum number of microsatellite loci or alleles for the construction of a stable dendrogram should be determined first. This will save time and labor for the development of core collections, especially for crops which have a large number of varieties or collections. In the present study we used SSRs to construct a phenetic dendogram of 14 cornerstone wheat breeding parents used prior to 1980, and 29 other well-known, large-scale planted varieties of Chinese wheat.

Materials and methods

Plant materials

Forty three Chinese common wheat varieties (Table 1) were employed, including 14 corner-stone breeding parents and 29 other large-scale planted varieties (Jing and Liu 1964; Jing et al. 1983, 1986, 1997; Zheng 2001). All varieties are kept by ICGR, CAAS (Chinese Academy of Agricultural Sciencies). Most of the cornerstone breeding parents, such as Funo, Mentana, Abbodanza and Orofen etc., are currently used widely in Chinese wheat production as well.

DNA extraction

Crude DNA was extracted from seeds. Seeds were crushed in mortars and homogenized in micro-tubes with extraction buffer (600 µl) containing 100 mM of Tris (pH 8), 50 mM of EDTA (pH 8), 500 mM of NaCl and 10 mM of beta-mercapoethanol. The

mixture was incubated at 65 °C for 30 min, then centrifuged at 14,000 rpm for 15 min. The supernatant was transferred into the same volume of Benzene-chloroform (24:1) and softly homogenized, then centrifuged at 14,000 rpm for 5 min. The supernatant was transferred into 2 volume of absolute ethanol (–20 °C) and a 1/10 volume of 10 M ammonium-acetate, incubated at 4 °C for 30 min and centrifuged at 14,000 rpm for 5 min. Precipitated DNA was air-dried, then re-suspended in 50 µl of TE and incubated with 10 mg/ml of RNAse at 37 °C for 15 min. DNA concentration was measured, and DNA was diluted to a suitable concentration.

PCR amplification and electrophoresis

A total of 126 WMS markers were selected from the microsatellite map published by Röder et al. (1995). On each chromosome, six loci were chosen, two on the terminal regions, two in the middle, and two close to the centromere. Three loci were selected on each arm. PCR amplification was performed in a 25-µl reaction solution containing 20 ng of genomic DNA, 250 nM of each primer, 200 µM of dNTPs, and 1 U of *Taq* polymerase. Thirty two PCR cycles were performed in a MJ PIC-200ZZZ; with each cycle consisting of denaturation at 95 °C for 30 s, annealing at either 50, 55 or 60 °C for 30 s, and extension at 72 °C for 1 min. The first cycle was preceded by a 5-min denaturation at 95 °C, and the last cycle was followed by a 10-min final extension at 72 °C. Denatured PCR products were separated in the 6% polyacrylamide gel in $1 \times$ TBE buffer for 50 min and stained with the silver method.

Data analyses

Microsatellite profiles were scored reflecting either the presence (1) or absence (0) of clear bands. Genetic similarities were estimated using the DICE coefficient, 2a/(2a+b+c), where "a" refers to alleles shared between two varieties, and "b" and "c" refer to alleles present in either one of the two varieties compared (Rohlf 1993). Eight dendrograms were clustered using the un-weighted pair group method average (UPGMA) using NTSYS-pc software (version 1.8, Rohlf 1993): one was based on the total 501 alleles; three of them were based on alleles located on the **A**, **B** and **D** genomes, respectively; and the other four were based on randomly selected 450, 400, 350 and 300 alleles. The MXCOMP procedure of the NTSYS-pc software was applied to determine the correlation between pairs of the DICE coefficient matrices estimated with 501, 450, 400, 350 and 300 alleles, and also used to test the correlation between the DICE coefficient matrices of the **A**, **B** and **D** genomes with the global matrix of 501 alleles.

Genetic diversity of each locus was expressed by allelic rich-ness, $\sum A_{ij}$, and genetic dispersion (Simpson index h_i), $h_i = 1 \cdot \sum_{j=1}^{i} p_{ij}^2$, where P_{ij} means the frequency of each allele at this locus. A diversity index (H_t) for each genome was computed by $H_t = \sum_{i=1}^{k} h_i/k,$, where k means the total number of loci studied within each genome, except those amplified in two or three genomes at the same time (Nei 1973).

Results

Levels of polymorphism detected

Out of 126 WMS markers, only 90 pairs of primers amplified stable bands. These were used to investigate the allelic diversity within varieties. A total of 501 alleles was obtained, with 166 of them belonging to loci within the **A** genome, and 174 and 161 belonging to loci within the **B** and **D** genomes respectively. In the three genomes, the average allelic richness was 4.7, 5.2 and 4.9 respectively. The diversity index values (H_t) of the **A**, **B** and **D** genomes were 0.53, 0.64 and 0.62, respectively.

Dendrograms of different numbers of alleles and different genomes

Eight dendrograms were constructed based on the genetic similarity shown in the clustering analysis. The dendrogram was constructed with 501 alleles (Fig. 1a), which divided all varieties, except Youzimai, into seven major groups. The dendrogram with 350 alleles clustered six Chinese landraces, Chinese Spring, Chengduguangtou, Mazamai, Jiangdongmen, Yada 1817 and Youzimai, into one large group, which was distinctly separated from the modern varieties (Fig. 1b). The relative positions of the various varieties on the trees were correlated with their pedigrees for most of the tested varieties (Fig. 1a, b). For example, Funo and its selections and most of its hybrid decadents, such as Bonong 7023, Yangmai1, Yangmai 5 and Yangmai 158, are clustered within a small group. Bainong 3217 and Fan 6, two decadents of Funo crossed by multi-parents, are a little bit far away from Funo. Orofen, a decadent of Mentana crossed by four other varieties are clustered one by one in the dendrogram. Jinan 13, a variety derived from Huixianhong/Abbondanza//Orofen, was clustered in the same small group with Orofen and Mentana (on details of the variety pedigrees, please see Jing et al. 1983, 1986, 1997; Jing and Zhuang 1996). However, the varieties in the second major group included Youbaomai, Bima 4, Beijing 8, Akagomughi and Fengchan 3, which exhibited no contribution from Strampelli's varieties. These varieties were located far from Funo, Villa Glori, Mentana, Orofen, Abbondanza, Ardito and their Chinese derivatives.

Large differences were detected between the three dendrograms based on SSR markers within the **A**-, **B**- or **D**- genomes independently. Several varieties were clustered in different groups in the three phenetic trees, such as Xiaoyan 6 and its male parent St2422-464. In the **A**-dendrogram, they were clustered side by side; however, in both the **B**- or **D**-dendrograms they are distinctly separated. Abbondanza and its two derivatives (Taishan 4 and Taishan 5) were clustered side by side in the 501 global tree and the tree based on 350 random alleles (Fig. 1). However, Taishan 5 was distinctly separated from Abbondanza and Taishan 4 in trees based on markers within either the **A**- or **D**-genomes. Five dendrograms were constructed based on 501, 450, 400, 350 and 300 random alleles. The number of alleles in each successive dendrogram did not bring any major change in groupings. Comparing the dendogram with 450 alleles and that of 501 alleles showed similar clustering in both dendrograms. Only three varieties (Youzimai, Fan 6 and St2422-464) changed their relative position in the 350-allele dendrogram compared to the 501-allele den-

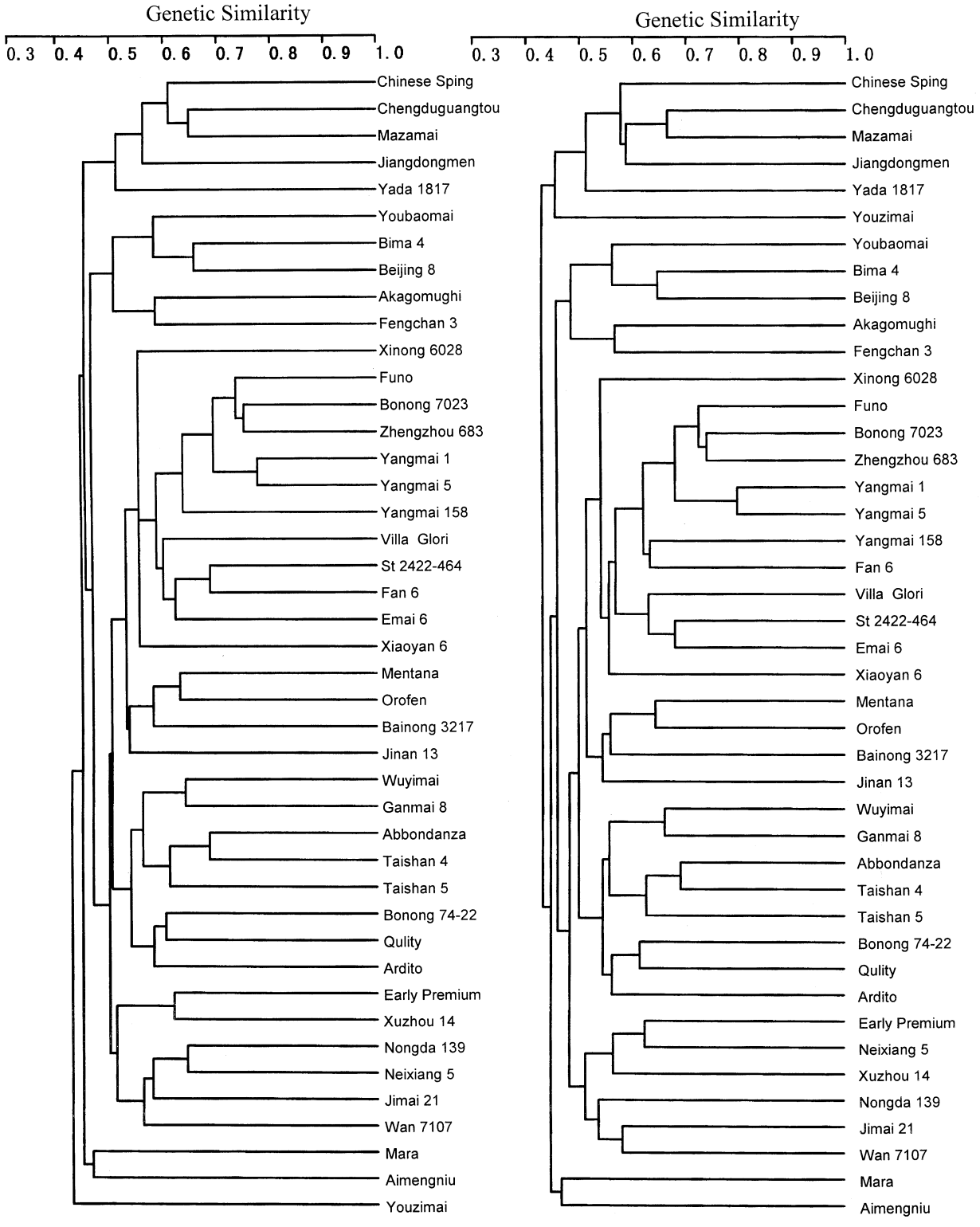

Fig. 1 The UPGMA dendrogram of Chinese corner-stone breeding parents and large-scale planted varieties based on 501 SSR alleles (**a**) and 350 random SSR alleles (**b**)

dogram (Fig. 1a, b). Youzimai exhibited the largest change, being separated from the Chinese landrace group in the global dendrogram. Fan 6 and St2422-464 clustered in the same major group in both dendrograms.

Correlation between genetic distance matrices

The correlation coefficients between genetic-distance matrices based on 501 and 450, 400, 350, 300, 250, 200 and 167 (one locus for each chromosome) random alleles

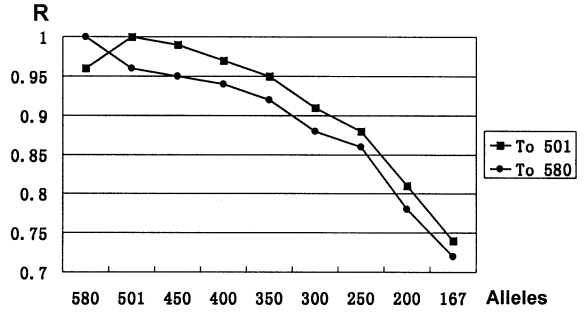

Fig. 2 The curve graph revealing the correlation trend of the genetic similarity matrix based on different numbers of alleles with those of 501 and 580 respectively

Table 2 Correlation coefficients of matrices based on different numbers of random alleles (167 to 580) compared to matrices based on 501 and 580 alleles

Alleles	580	501	450	400	350	300	250	200	167
501	0.96	1.0	0.99	0.97	0.95	0.91	0.88	0.81	0.74
580	1.0	0.96	0.95	0.94	0.92	0.88	0.86	0.78	0.72

are presented in Table 2, and the correlation trend is shown in Fig. 2. When the number of alleles used was raised to 580, the correlation coefficient between its genetic similarity matrix and that of 501 was 0.96. The corresponding index between the 580 and 400 alleles was 0.94 (Table 2). Therefore, the increase in the number of alleles did not significantly change the information offered by the cluster trees. The correlation coefficients between genetic distance matrices based on 501 alleles, and loci within the **A**, **B** and **D** genomes independently, separately were 0.77, 0.76 and 0.67, respectively. The reduced number of alleles may be the cause of these low coefficients. The correlation coefficients of allele combinations with **A+B** (340 alleles), **A+D** (327 alleles) and **B+D** (335 alleles) in the 501 genetic similarity matrix were estimated and were 0.91, 0.93 and 0.89. The low correlation coefficient of the third combination indicates that alleles should come from all of the three genomes instead of only one or two genomes, so that genetic relationships between varieties can be accurately evaluated.

Discussion

Results of our research showed that the genetic diversity of the three genomes is **B > D > A** in the 43 analyzed varieties, which was reflected by the average allelic richness of per locus and the average genetic dispersion index (*Ht*) of each genome. This is basically consistent with information of the genetic diversity revealed by RFLP markers (Jia et al. 2001). In addition, correlation coefficients between the genetic similarity matrix of 501 alleles, and those based on markers within every one of three genomes independently were 0.77, 0.76, and 0.67. These data and the very low correlation coefficient between the global matrix and that of the **B+D** genomes suggest that true genetic relationships in polyploid species should be estimated with markers in all genomes, and that the loci should cover all pairs of homologous chromosomes.

In the dendrogram tree of 350 alleles, all of the six Chinese common wheat landraces were clustered in one major group (Fig. 1b). In the 501 global tree they were clustered into one major group again except for Youzimai (Fig. 1a). This was contradictory to the traditional view that Chinese landraces have high genetic diversity based on morphological characters. Our results of SSRs agree with that from the HMW glutenin subunit component (*Glu-1*) analysis in 5,129 candidate core collections of Chinese wheat germplasms (Zhang et al. 2002b). The genetic diversity of Chinese landraces is still uncertain and should be further evaluated.

There are several limitations in using pedigree data as a tool to study genetic relationships. Many varieties, especially landraces and semi-wild relatives, have no clear pedigree data or have ambiguous ancestry. In European cultivated spelt wheat (*Triticum spelta* L.), a given genotype is not necessarily situated at the mid-distance from its two parents (Bertin et al. 2001). For example, in our study even though Bima 4 and Early Premium are the two parents of Beijing 8, Bima 4 and Beijing 8 clustered into the same group, Early Premium is distinctly separated from them. This suggests that each parent played different roles in the genetic construction of its offspring, probably due to gene drag and artificial selection. This also suggests that the same parents could produce quite different varieties in different selection programs. Examples of this include Early Premium and Mentana, and their hybrid Xuzhou14, as well as Wuyimai and Abbondanza and their hybrid Ganmai 8 (Fig. 1a, b). If the same parents can produce different offspring, an analysis using the coefficients of co-ancestry would lead to incorrect results. For example, Jianan 13 and Taishan 5 share the same parents, but they cluster into different groups (Fig. 1). This shows that varieties with the same coefficient of co-ancestry may have different genetic similarities than their parents. As a result, SSR data seems to more-accurately reflect genetic relationships than pedigree data.

Previous genotype research found that 113 alleles were enough to cluster cultivated spelt wheat in the European group (Bertin et al. 2001). Successful genotype identification was performed in soybean [*Glycine max* (L.) Merr.] with seven microsatellites (Rongwen et al. 1995). Genetic distance and dendrograms were generated for rice (*Oryza sativa* L.) using ten microsatellites (Garland et al. 1999), 15 in barley (*Hordeum vulgare* L.) (Struss and Plieske 1998), and 23 in common wheat (Plaschke et al. 1995). The locus number has been used to construct dendrograms, as well as to identify genotypes and compute genetic distances. However, we believe that the number of alleles might be more suitable than locus number to construct dendrograms. This is because different pairs of primers or loci have a different genetic diversity, which results in different allelic rich-

ness in the same group of samples. For example, several loci just have two alleles at each locus; the highest can reach 8–9 alleles in these varieties.

The use of only 167 alleles (one locus with high diversity for each homologous chromosome) allowed us to discriminate all Chinese common wheat genotypes in our study. The correlation coefficient of the genetic distance matrix between 501 alleles and 167 alleles was 0.74, which indicated that 167 alleles were sufficient to differentiate all varieties, even inbred selection lines such as Funo and its selection Wan 7107. However, 167 alleles are not enough to construct a stable dendrogram to objectively reflect genetic relationships. It is important to not only discriminate varieties but also to obtain a stable dendrogram that reflects the true genetic relationships among wheat varieties so that core collections can be established. No references are available that describe the "saturation point" for constructing a stable dendrogram, or developing core collections with the least number of alleles in common wheat; our study (Fig. 1a, b) suggests that 350 to 400 alleles are required to objectively construct a dendrogram tree for common wheat (Fig. 1a, b; Table 2, Fig. 2). Finding such a saturation point for crops such as wheat, that have a large number of accessions, will minimize the time and costs in establishing a core collection. Further studies on the relationship between the allelic number needed and the sample size in completely random samples are now in progress in our group.

Acknowledgements We are grateful to Dr. D. Johnson of USDA-ARS, FRRL and Dr. M. Carmen de Vicente of IPGRI for critically reading and discussing the manuscript. This research was financial supported by the Chinese Ministry of Sciences and Technology (project number G1998010202 to XYZ).

References

Beckmann JS, Soller M (1990) Toward a unified approach to genetic mapping of eukaryotes based on sequence tagged microsatellite sites. Bio/Technology 8:30–32

Bertin P, Grégoire D, Massart S, Frodimont D de (2001) Genetic diversity among European cultivated spelt revealed by microsatellites. Theor Appl Genet 102:148–156

Danin-Poleg Y, Reis N, Baudracco-Arnas S, Pitrat M, Staub JE, Oliver M, Arus P, deVicente CM, Katzir N (2000) Simple sequence repeats in *Cucumis* mapping and map merging. Genome 43:963–974

Garland SH, Lewin L, Abedinia M, Henry R, Blakeney A (1999) The use of microsatellite polymorphisms for the identification of Australian breeding lines of rice (*Oryza sativa* L.). Euphytica 108:53–63

Hammda H, Petrino MG, Kakunaga T (1982) A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. Proc Natl Acad Sci USA 79:6465–6469

Jia JZ, Zhang ZB, Devos K, Gale MD (2001) Genetic diversity in wheat revealed by restriction fragment length polymorphism (RFLP) markers. (in Chinese). Science in China (Series C) 31:13–21

Jing SB, Liu DA (chief editors) (1964) Flora of Chinese Wheat Varieties (in Chinese). China Agricultural Publishing House, Beijing

Jing SB, Zhuang QS, Wu ZS, Huang PM, Bo YJ (1983) Chinese Wheat Varieties and their genealogics (in Chinese). China Agricultural Publishing House, Beijing

Jing SB, Zhuang QS, Sha ZG, Wu ZS, Cheng HY, Lu MZ, Zhao DF, Dong YS, Cai X (chief editors) (1986) Flora of Chinese wheat varieties from 1962–1982 (in Chinese). China Agricultural Publishing House, Beijing

Jing SB, Dong YS, Qian MM, Zheng DS (chief editors) (1997) Flora of Chinese Wheat Varieties from 1983–1993. (in Chinese). China Agricultural Publishing House, Beijing

Kihara H (1944) Discovery of the DD-analyzer, one of the ancestors of *vulgare* wheat. Agri Hort 19:889–890

McFadden ES, Sears ER (1946) The origin of *Triticum spelta* and its free-threshing hexaploid relatives. J Hered 37:107–116

Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci USA 70:3321–3323

Pestsova E, Ganal MW, Röder MS (2000) Isolation and mapping of microsatellite markers specified for the D genome of bread wheat. Genome 43:689–697

Plaschke J, Ganal MW, Röder MS (1995) Detection of genetic diversity in closely related bread wheat using microsatellite markers. Theor Appl Genet 91:1001–1007

Röder MS, Plaschke J, König SU, Börner A, Sorrells ME, Tanksley SD, Ganal MW (1995) Abundance, variability and chromosomal location of microsatellites in wheat. Mol Gen Genet 264:327–333

Röder MS, Korzun V, Wendehake K, Plaschke J, Tixier MH, Leroy P, Ganal MW (1998) A microsatellite map of wheat. Genetics 149:2007–2023

Rohlf FJ (1993) NTSYS-pc: numerical taxonomy and multivariate analysis system. Applied Biostatics Incorporated, New York

Rongwen J, Akkaya MS, Bhagwat AA, Lavi U, Cregan PB (1995) The use of microsatellite DNA markers for soybean genotype identification. Theor Appl Genet 90:43–48

Struss D, Plieske J (1998) The use of microsatellite markers for detection of genetic diversity in barley populations. Theor Appl Genet 97:308–315

Zhang XY, Dong YS, You GX, Wang LF, Li P, Jia JZ (2002a) Allelic variation of *Glu-A1 Glu-B-1* and *Glu-D1* in Chinese released wheat varieties in the last 50 years. Agric Sci in China 1:36–44

Zhang XY, Pang BS, You GX, Wang LF, Jia JZ, Dong YS (2002b) Allelic variation and genetic diversity at *Glu-1* loci in Chinese wheat germplasms (in Chinese). Sci Agric Sinica (in press)

Zheng DS (2001) Analysis on wheat varieties with an annual planting area larger than 667,000 hm^2 in China (in Chinese). J Plant Genet Resources 2:49–55