Theor Appl Genet (2004) 108:920–930
DOI 10.1007/s00122-003-1502-y

ORIGINAL PAPER

V. Roussel · J. Koenig · M. Beckert · F. Balfourier

Molecular diversity in French bread wheat accessions related to temporal trends and breeding programmes

Received: 17 July 2003 / Accepted: 9 October 2003 / Published online: 12 November 2003

Abstract A set of 41 wheat microsatellite markers (WMS), giving 42 polymorphic loci (two loci on each chromosome), was used to describe genetic diversity in a sample of 559 French bread wheat accessions (landraces and registered varieties) cultivated between 1800 and 2000. A total of 609 alleles were detected. Allele number per locus ranged from 3 to 28, with a mean allele number of 14.5. On the average, about 72% of the total number of alleles were observed with a frequency of less than 5% and were considered to be rare alleles. WMS markers used showed different levels of gene diversity: the highest PIC value occurred in the B genome (0.686) compared to 0.641 and 0.659 for the A and D genomes, respectively. When comparing landraces with registered varieties gathered in seven temporal groups, a cluster analysis based on an F_{st} matrix provided a clear separation of landraces from the seven variety groups, while a shift was observed between varieties registered before and after 1970. There was a decrease of about 25% in allelic richness between landraces and varieties. In contrast, when considering only registered varieties, changes in diversity related to temporal trends appeared more qualitative than quantitative, except at the end of the 1960s, when a bottleneck might have occurred. New varieties appear to be increasingly similar to each other in relation to allelic composition, while differences between landraces are more and more pronounced over time. Finally, considering a sub-sample of 193 varieties representative of breeding material selected during the twentieth century by the six most important plant breeding companies, few differences in diversity were observed between the different breeding programmes. The observed structure of diversity in French bread wheat collections is discussed in terms of consequences, both for plant breeders and for managers of crop genetic resources.

Communicated by H.H. Geiger

V. Roussel · J. Koenig · M. Beckert · F. Balfourier (✉)
Amélioration et Santé des Plantes (UMR 1095),
INRA, 234 avenue du Brézet,
63039 Clermont-Ferrand, Cedex 2, France
e-mail: balfour@clermont.inra.fr
Tel.: +33-473-624346
Fax: +33-473-624453

Introduction

Wheat is one of the most important cultivated crops in the world with a worldwide production of 563.2 MT in 2002 (39 MT in France) (http://wbc.agr.state.mt.us/prodfacts/wf/wptwp.html). Landraces were mainly cultivated in France until the middle of the nineteenth century and then gradually replaced by breeding lines (Bustarret 1944). One of the first plant-breeding programmes was developed by Vilmorin during the second part of the nineteenth century. This breeder developed the first famous varieties released in France at the beginning of the twentieth century (i.e. Dattel, Massy, Tresor, Hatif-Inversable, Bon-Fermier, Allies, Paix, etc.). Many different breeding companies such as Schribaux, Blondeau, Benoist, Tourneur, Desprez and Lepeuple then developed new cultivars, essentially by crossing Vilmorin's genotypes. Finally, after the Second World War, both the use of foreign germplasm in plant-breeding programmes and the introduction of dwarf genes produced the best elite lines developed during the last decades by an increasing number of breeding companies.

Intensive plant breeding is generally considered to be a practice that leads to reduced genetic diversity. This reduction can induce modifications in gene frequencies and, consequently, changes in the future adaptability of the crop. Thus, it seems necessary to conserve genetic variation for the improvement of future varieties. That is why significant research on genetic resource characterisation was done on major cereal crops such as rice (Li et al. 2002; Ni et al. 2002; Gao et al. 2003), barley (Russel et al. 2000; Liu et al. 2002; Matus et al. 2002; Baek et al. 2003; Koebner et al. 2003; Lund et al. 2003; Sjakste et al. 2003) and maize (Dillmann et al. 1997), for example. Efforts have also been made to characterise the genetic

diversity in wheat using morphological traits (DeLacy et al. 2000), isozymes (Guadagnuolo et al. 2001), storage protein diversity (Branlard et al. 2001) and molecular markers such as RAPD (Chandrashekhar et al. 1993; Cao et al. 2000), AFLP (Donini et al. 2000; Manifesto et al. 2001), RFLP (Jones et al. 1997; Kim et al. 1997; Ward et al. 1998) or microsatellites (Plaschke et al. 1995; Prasad et al. 2000; Stachel et al. 2000; Ben Amer et al. 2001; Christiansen et al. 2002). One of the most promising molecular markers seems to be microsatellites which appear very useful for comparisons of closely related accessions (Guadagnuolo et al. 2001). Microsatellites are short tandem repeats of DNA sequences. The advantage of such markers is their abundance in the genome, their high polymorphism levels, their co-dominance and their high reproducibility. They can be used to determine genetic relationships between crop varieties. Moreover, their location on the genome map is already known for most of them.

In this paper, we report the results of a study of 559 French bread wheat accessions cultivated between 1800 and 2000, using 41 microsatellite primers. The purposes of this study were (1) to provide a deep insight into the genetic diversity of the French bread wheat collection, (2) to compare the level and the distribution of genetic variability between groups of varieties according to their registration year and to their breeding origins, and (3) to assess the potential application of our results for further studies on the evaluation and conservation of wheat genetic resources. The findings are discussed in the context of historical and current French wheat breeding procedures.

Materials and methods

Plant material and DNA extraction

Seeds from 559 genotypes of *Triticum aestivum,* obtained from the Centre of Biological Resources on Cereal Crops of Clermont-Ferrand (INRA), were used for this study. These genotypes were old French landraces from the nineteenth century and varieties selected between 1840 and 2000 (the list of this material and passport data are available upon request). First, we decided to organise accessions into eight groups according to their period of registration (varieties) or cultivation (landraces). Table 1 lists the number of varieties in each decadal group. Next, varieties were considered in terms of breeding programmes. Sixty-four different breeding companies produced the 559 genotypes. In order to compare groups with a sufficient number of accessions during a relatively similar production period, we chose to analyse only six breeding companies which produced more than 18 genotypes in our collection during the last eight decades of the twentieth century. Table 2 lists the number of accessions in each of these six breeding groups (total = 193 accessions).

Table 1 Registration period and number of investigated accessions

Decadal group	Number of accessions
Landraces	62
1840–1929	60
1930–1944	62
1945–1959	50
1960–1969	55
1970–1979	62
1980–1989	109
1990–2000	99

Table 2 Name, number of varieties and production period of the six groups

Breeder	No. of accessions	Production period
Benoist	38	1936–1998
Blondeau	23	1936–1995
Desprez	53	1932–1998
INRA	25	1935–1998
Tourneur	19	1924–1986
Verneuil-Vilmorin	35	1923–1997

All the seeds used for DNA extraction were obtained from self-pollinated ears. Five to six plants per accession were pooled and bulk genomic DNA was isolated using a SIGMA GenElute Plant Genomic DNA Kit (G2N-350).

Microsatellite analysis

A set of 41 microsatellite primer pairs, giving 42 polymorphic loci, were selected according to their location on chromosome arms, their readability and their reproducibility. All these microsatellites were previously developed and mapped by Röder et al. (1998) (gwm markers from IPK Gatersleben), except for one (cfd71) originating from INRA-Clermont-Ferrand (Guyomarc'h et al. 2003), which detects two loci. Each of the 42 loci was located on one different chromosome arm, in order to cover the whole bread wheat genome.

PCR reaction was performed according to Röder et al. (1998): an initial denaturation (3 min at 94°C), and 35 cycles of 30 s at 94°C for denaturation, 30 s at 50°C (between 45°C and 60°C, depending on the primer) for annealing and 30 s at 72°C for extension, followed by a final extension step of 5 min at 72°C. Fragment analysis was carried out on an automated sequencer ABI Prism 377 from Perkin Elmer, and fragment sizes were calculated using the Genescan Analysis Software (version 3.1) and Genotyper (version 2.5) computer packages, where different alleles were represented by different amplification sizes of tandem repeat.

For each microsatellite locus and for the whole set of accessions, the total number of alleles observed and the number of rare alleles were recorded. An allele was considered to be rare if its frequency was lower than 0.05. Finally, the polymorphic information content (PIC) value was calculated for each locus according to the formula:

$$\mathrm{PIC} = 1 - \sum p_i^2 \qquad (1)$$

where p_i is the frequency of the ith allele. This value, which is equivalent to the genetic diversity index (H) of Nei, provides an estimate of the discriminatory power of a microsatellite locus.

Diversity analysis

To characterise the genetic diversity of the different accession groups and identify differences between them, diversity parameters such as allelic richness (A) and Nei's average gene diversity (H_e) were calculated for each group. However, since the sample size of each group was different, we used the allele rarefaction method (Petit et al. 1998) to estimate A and H_e for comparable group sample size. The rarefaction method is used to standardise the allelic richness across populations. This technique makes it possible to evaluate the expected number of different alleles among equal-sized samples drawn from several different populations. We chose a basic group sample size of 39 for decadal groups and 17 for

breeding groups. $a_1, \ldots, a_k$ are the different alleles at a single locus and $N_1, \ldots, N_k$ are the number of times they appeared in the N-size sample of each group. Therefore, the expectation of the number of alleles in a sample of size g ($g \leq N$) is:

$$\hat{r}_{(g)} = \sum_i \left[1 - \left(\frac{C^g_{N-Ni}}{C^g_N}\right)\right] \quad (2)$$

This method gives an estimate of the total number of alleles which could have been obtained with a chosen effective size 'g'. Thus, in order to have an accurate estimation of the rare allele number for each group, we calculated the ratio r = total allele number after/total allele number before rarefaction for each of them. We then multiplied the initial number of rare alleles by this ratio.

A sign test (Dagnélie. 1986) was used to detect significant differences in allelic richness between groups. The null hypothesis of this test is:

$$H_0 : P(+) = P(-) = \frac{1}{2} \quad (3)$$

with: $P(+)$ = probability of observing a positive difference and $P(-)$ = probability of observing a negative difference.

Next, the probability of obtaining the same number of positive and negative differences as observed is calculated with:

$$P(X \leqslant x) = \left(\frac{1}{2}\right)^n \sum_{i=0}^{x} C_n^i \quad (4)$$

H_0 is rejected if $P \leq 0.05$.

A study of rare alleles was also done for temporal groups. This study was performed on results before allele rarefaction method. A list of all the alleles detected was drawn up and the number of accessions containing these alleles was found. We were only interested in the "presence/absence" of rare alleles in each group. A table was obtained showing all the alleles for all groups, where 'O' indicates that there are no accessions containing this allele, and X indicates that at least one accession contains this allele.

First, the differences were calculated between landraces and all other groups. Then, the same procedure was used between each pair of successive groups on a time scale. We considered a difference to be an allele which was not present in the first group and which appeared in the second group, or an allele which was present in the first group and not in the second one. So, the number (and percentage) of alleles appearing and disappearing between two compared groups was counted, and a graph was therefore established, showing the percentage of rare alleles appearing or disappearing between groups, with 100% representing the total number of rare alleles found.

Population structure analysis

To compute the analysis of molecular variance (AMOVA) among and within groups, data were analysed using Arlequin software (Schneider et al. 1996).

In order to specify the level of diversity between groups, we also calculated F_{st} and R_{st} (Slatkin 1995) values, using the whole set of microsatellite loci:

$$F_{st} = \frac{\hat{\sigma}^2\text{inter}}{\hat{\sigma}^2\text{inter} + \hat{\sigma}^2\text{intra}} \quad (5)$$

with:

$$\hat{\sigma}^2\text{inter} \quad (6)$$

= variance of allelic frequencies between groups.

$$\hat{\sigma}^2\text{intra} \quad (7)$$

= variance of allelic frequencies within a group.

$$R_{st} = \frac{\hat{\sigma}^2\text{inter}}{\hat{\sigma}^2\text{inter} + \hat{\sigma}^2\text{intra}} \quad (8)$$

with:

$$\hat{\sigma}^2\text{inter} \quad (9)$$

= variance of allele sizes between groups.

$$\hat{\sigma}^2\text{intra} \quad (10)$$

= variance of allele sizes within a group.

F_{st} is generally estimated under an infinite allele model (IAM) mutation model hypothesis for microsatellites, while R_{st} is estimated under a stepwise mutation model (SMM) one (Ohta and Kimura. 1973; Di Rienzo et al. 1994). As these mutation models were unknown for the microsatellite used, the two parameters, F_{st} and R_{st}, were computed for each possible pair of groups and compared. Cluster analyses were then performed on the different F_{st} and R_{st} matrices using an UPGMA algorithm.

Finally, the genetic distance between each pair of the 559 accessions was calculated by 1-P, where P is the proportion of shared alleles for the 42 loci (Nei and Li 1979). The accessions were then scaled in a two-dimensional space using non-metric scaling (Kruskal et al. 1978). This is a type of method for estimating the coordinates of a set of objects in a space of specified dimensionality on the basis of data that measure the distances between a pair of objects.

Results

Microsatellite analysis and allelic distribution

Chromosomal location, the total number of alleles, the number of rare alleles and PIC values are presented in Table 3 for each microsatellite locus, while results are averaged by genome and chromosome groups in Table 4. Different alleles represent different sizes of amplification products. Seeds from self-pollinated plants were used for DNA extraction and most amplification products revealed only one allele per locus. When two different alleles were present at a single locus, data were considered to be missing, since it was not possible to distinguish between DNA contamination and heterozygosity. In the same way, since it was difficult to distinguish between null alleles and non-amplification due to experimental errors, null alleles were not recorded except for one locus (*Xgwm135*), where gel interpretation was quite evident. Under these conditions, less than 5% of the data were considered to be missing and a total of 609 alleles were detected from the 42 amplified loci. The total number of alleles per locus ranged from 3 for locus *Xgwm664* to 28 for locus *Xgwm539*, with an average number of 14.5. The number of rare alleles was also variable and ranged from 2 for *Xgwm664* and *Xgwm272* to 23 for *Xgwm539*. On the average, about 72% of the total number of alleles was considered to be rare alleles.

The microsatellite markers used show different levels of gene diversity. PIC values ranged from 0.214 to 0.868 with an average of 0.662 for all markers. The highest PIC value occurred in the B genome with 0.686, compared to 0.641 and 0.659 for the A and D genomes, respectively. There was a high correlation coefficient between gene diversity and the number of alleles (r=0.69). If we

Table 3 Chromosomal location, total number of alleles, number of rare alleles and polymorphic information content (PIC) values for 42 microsatellite loci

Location	Locus	Alleles	Rare alleles	PIC	Location	Locus	Alleles	Rare alleles	PIC	Location	Locus	Alleles	Rare alleles	PIC
1A	*Xgwm99*	11	7	0.495	1B	*Xgwm11*	16	11	0.731	1D	*Xgwm642*	6	4	0.449
1A	*Xgwm135*	19	17	0.454	1B	*Xgwm413*	15	11	0.711	1D	*Xgwm337*	17	12	0.762
2A	*Xgwm312*	22	17	0.761	2B	*Xgwm257*	5	3	0.567	2D	*Xgwm539*	28	23	0.833
2A	*Xgwm372*	22	17	0.805	2B	*Xgwm120*	17	13	0.754	2D	*Xgwm261*	14	10	0.553
3A	*Xgwm5*	17	12	0.795	3B	*Xgwm181*	16	12	0.659	3D	*Xgwm664*	3	2	0.118
3A	*Xgwm480*	11	10	0.215	3B	*Xgwm566*	8	5	0.676	3D	*Xgwm161*	13	8	0.793
4A	*Xgwm610*	10	7	0.502	4B	*Xgwm251*	16	11	0.835	4D	*Cfd71D*	18	14	0.778
4A	*Cfd71A*	7	5	0.386	4B	*Xgwm149*	9	6	0.440	4D	*Xgwm194*	19	11	0.868
5A	*Xgwm415*	8	5	0.641	5B	*Xgwm408*	18	15	0.710	5D	*Xgwm272*	6	2	0.612
5A	*Xgwm186*	15	9	0.839	5B	*Xgwm234*	12	5	0.848	5D	*Xgwm190*	15	11	0.714
6A	*Xgwm169*	25	19	0.856	6B	*Xgwm219*	23	17	0.850	6D	*Xgwm469*	17	13	0.776
6A	*Xgwm427*	18	12	0.811	6B	*Xgwm626*	5	3	0.387	6D	*Xgwm325*	10	6	0.606
7A	*Xgwm260*	18	14	0.743	7B	*Xgwm400*	17	13	0.685	7D	*Xgwm44*	16	11	0.725
7A	*Xgwm233*	10	6	0.678	7B	*Xgwm46*	17	13	0.755	7D	*Xgwm437*	20	17	0.645
Total group A		213	157		Total group B		194	138		Total group D		202	144	

Table 4 Mean values of total allele number, rare allele number and PIC values for each homologous group of chromosome and genome

Homologous group	Alleles	Rare alleles	Rare allele (%)	PIC
1	14.00	10.33	73.81	0.60
2	18.00	13.83	76.85	0.71
3	11.33	8.17	72.06	0.54
4	13.17	9.00	68.35	0.54
5	12.33	7.83	63.51	0.73
6	16.33	11.67	71.43	0.71
7	16.33	12.33	75.51	0.71
A	15.21	11.21	73.71	0.64
B	13.86	9.86	71.13	0.69
D	14.43	10.29	71.29	0.66

compare the seven homologous groups, we can see that the highest PIC values occurred in chromosomes 5, 6 and 2, and the lowest were observed for chromosome 4 (Table 4).

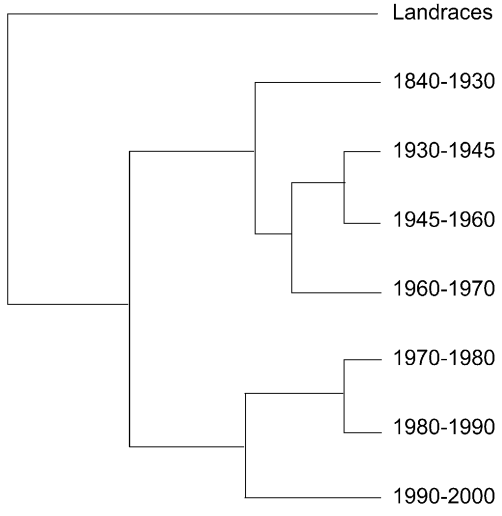

Fig. 1 Cluster analysis of eight decadal groups, using an F_{st} parameter and 42 microsatellite loci

Temporal variation in diversity

Considering the whole set of 42 microsatellites, F_{st} and R_{st} matrices among group values are presented in Table 5. The results of structuration analysis are quite similar using the F_{st} or R_{st} parameters; the correlation coefficient between the two parameters is very high (r=0.98). The dendrogram resulting from the cluster analysis based on the F_{st} parameter is presented on Fig. 1: landraces are separate while decadal groups form two clusters, according to a period of time—before and after the 1970s.

To analyse the separation between landraces and varieties, we first applied an AMOVA (Table 6). The between-variance component only represents 5.83% of the whole variation.

Table 5 Matrix of F_{st} (above) and R_{st} (below) values between the eight decadal groups, calculated for 42 microsatellite loci

	Landraces	1840–1930	1930–1945	1945–1960	1960–1970	1970–1980	1980–1990	1990–2000
Landraces	0	0.0091	0.0184	0.0187	0.0239	0.0289	0.0262	0.0268
1840–1930	0.0087	0	0.0047	0.0067	0.0107	0.0197	0.0188	0.0223
1930–1945	0.0182	0.0058	0	0.0021	0.0072	0.017	0.0167	0.0187
1985–1960	0.0119	0.0034	0.0000	0	0.0033	0.0105	0.0142	0.0154
1960–1970	0.0202	0.0086	0.0036	0.0012	0	0.0035	0.007	0.0121
1970–1980	0.0286	0.0147	0.0108	0.0067	0.0008	0	0.0022	0.0106
1980–1990	0.0251	0.0141	0.0127	0.009	0.0037	0.0007	0	0.0053
1990–2000	0.0177	0.016	0.0151	0.0114	0.01	0.0099	0.004	0

Table 6 Analysis of molecular variance (AMOVA): landraces versus varieties

Source	*df*	Variance component	Variation accounted for (%)
Between	1	0.79***	5.83
Within	557	12.81***	94.17
Total	558	13.60	100.00

*** significant at P<0.001

Table 7 AMOVA: landraces vs varieties before the 1970s vs varieties after the 1970s

Source of variation	*df*	Variance component	Variation accounted for (%)
Among temporal groups	2	0.43** (V_a)	3.23
Among decadal group	5	0.30*** (V_b)	2.25
Within temporal group			
Within decadal group	551	12.46*** (V_c)	94.51
Total	558	13.19	

***, ** significant at P<0.001 and P<0.01, respectively

Table 8 Sign test on differences between decadal groups for allelic richness. For each pair-wise comparison, the number of positive / negative differences and the probability to obtain the same number are given

	Landraces	1840–1930	1930–1945	1945–1960	1960–1970	1970–1980	1980–1990	1990–2000
Landraces	0	–	–	–	–	–	–	–
1840–1930	30+/11- *P*<0.001	0	–	–	–	–	–	–
1930–1945	35+/6- *P*<0.001	17+/24- *P*<0.068	0	–	–	-	–	–
1945–1960	31+/10- *P*<0.001	29+/12- *P*<0.003	10+/31- *P*<0.001	0	–	–	–	–
1960–1970	38+/3- *P*<0.001	32+/9- *P*<0.01	25+/16- *P*<0.047	35+/6- *P*<0.001	0	–	–	–
1970–1980	34+/7- *P*<0.001	29+/12- *P*<0.003	25+/16- *P*<0.047	36+/5- *P*<0.001	15+/25- *P*<0.037	0	–	–
1980–1990	36+/5- *P*<0.001	29+/12- *P*<0.003	22+/19- *P*<0.11	34+/7- *P*<0.001	15+/26- *P*<0.028	21+/20- *P*<0.12	0	–
1990–2000	34+/7- *P*<0.001	27+/14- *P*<0.016	20+/21- *P*<0.12	33+/8- *P*<0.001	12+/29- *P*<0.001	20+/21- *P*<0.12	19+/22- *P*<0.11	0

In order to enforce the structuration analysis as a function of a time scale, AMOVA was then conducted on three temporal groupings of accessions: landraces, varieties registered before 1970 (i.e. four decadal groups: 1840–1930, 1930–1945, 1945–1960 and 1960–1970), and after 1970 (three decadal groups: 1970–1980, 1980–1990 and 1990–2000). The results (Table 7) clearly show that the within-decadal group component (V_c) of the molecular variance is overwhelmingly dominant. Although V_a and V_b components are significant, only a minor part of the overall variance appears to be the result of any temporal drift between decadal groups.

Finally, to characterise the genetic diversity of the eight different groups and observe the evolution of the diversity on a time scale, total allelic richness was estimated on sample of comparable size, using Petit's rarefaction allele method (1998). Figure 2 shows the allelic richness of landraces and each successive decadal group. The results of the sign test on allele number differences are shown in Table 8 for each pair-wise comparison. Although this test is not very powerful, it

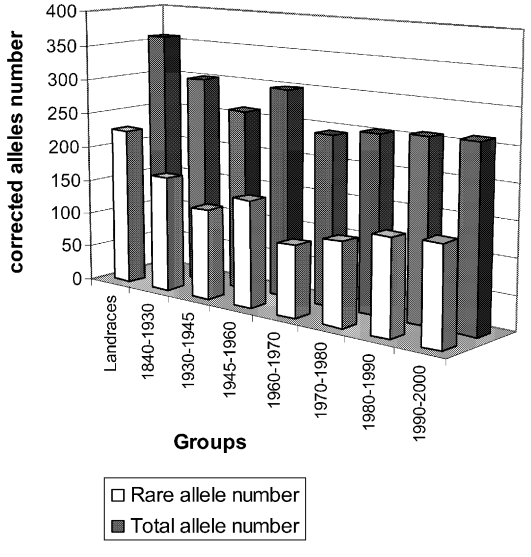

Fig. 2 Corrected allele number for eight decadal groups

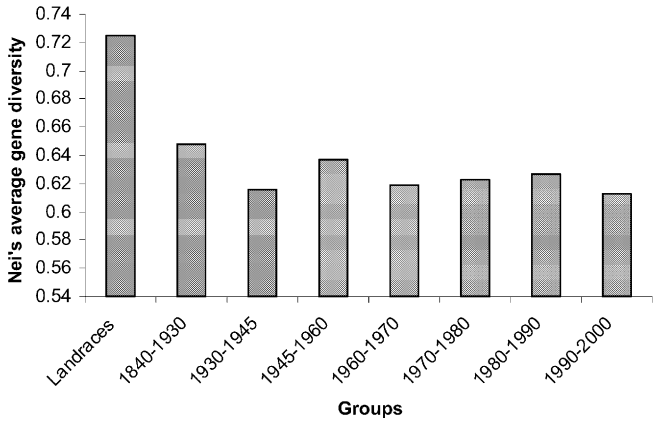

Fig. 3 Nei's average gene diversity for eight decadal groups

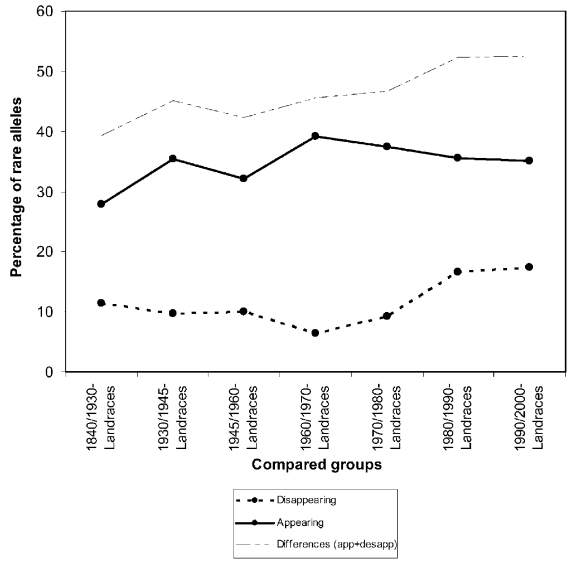

Fig. 4 Percentage of rare alleles appearing or disappearing in decadal groups, compared to landraces

gives a relatively good idea of the significant differences between groups. A decrease in allelic richness between landrace period and all the other decadal periods is always significant, since allelic richness regularly decreases from the beginning of the selection period (1840) until the end of the 1960s, except during a short period (1945–1960) where it increases. After the 1960s and until the end of the twentieth century, we observe a second phase where allelic richness is restored at a level corresponding to 75% of the initial diversity present in landraces. As for the Nei's average gene diversity (Fig. 3), we observe a quite similar picture during the same period.

In order to distinguish qualitative variations in allelic diversity, the number of rare alleles was also estimated in each decadal group; the most important number of rare alleles appears in the landrace group. Figure 4 shows the variation in the percentage of rare alleles observed between landraces and all of the other groups. We chose

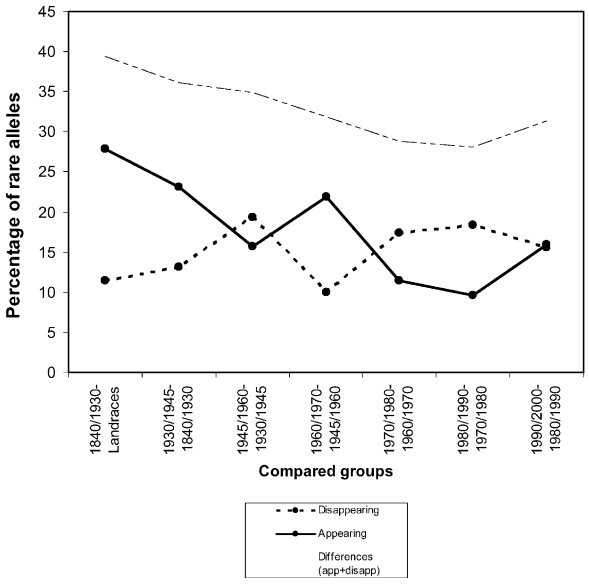

Fig. 5 Evolution of percentage of rare alleles appearing or disappearing, in successive decadal groups

to study only rare alleles because they are the only ones which can vary between groups. The lowest curve represents the percentage of rare alleles disappearing between landraces and a compared group, while the medium one represents the percentage of rare alleles appearing between landraces and the same compared group. These two curves are symmetric: when the number of disappearing alleles increases, the number of appearing ones decreases, and conversely. However, when we add the two categories of alleles (highest line), we observe that there is a regular increase of differentiation over a period of time in rare allele numbers between landraces and the different decadal groups; in other words, the more recent the varieties, the more different their rare allele composition is, compared with landraces.

To look at the evolution over a period of time, Fig. 5 shows the same results of variation in the percentage of rare alleles between successive decadal groups. We can also observe symmetric curves between appearing and disappearing rare alleles. Between 1840 and 1930, there are more rare alleles appearing than rare alleles disappearing, and there is an inversion of this phenomenon after the 1960s, where we can observe a greater number of rare alleles disappearing. When the two categories of alleles are added (highest line), we obtain a curve with a negative slope, indicating that the global differentiation is slower in recent times, compared to the oldest decades. In other words, the more recent the varieties, the more similar they are to each other in relation to their rare allele composition.

Table 9 AMOVA: breeder group effect

Source	*df*	Variance component	Variation accounted for (%)
Between breeder group	5	0.56***	4.43
Within breeder group	187	12.02***	95.57
Total	192	12.58	100.00

*** significant at P<0.001

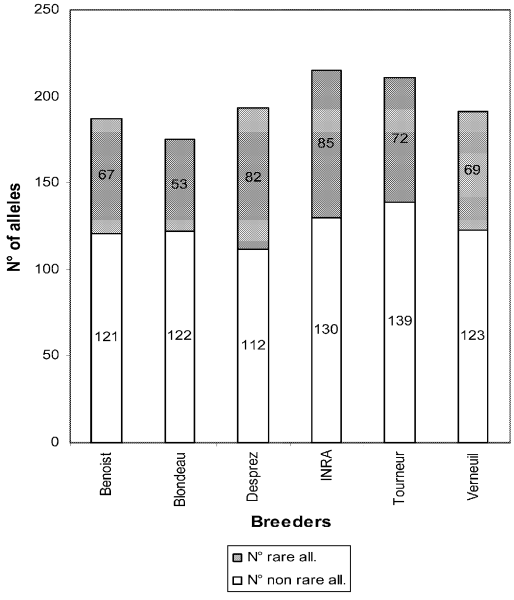

Fig. 6 Differences in allele number between breeders for common and rare alleles

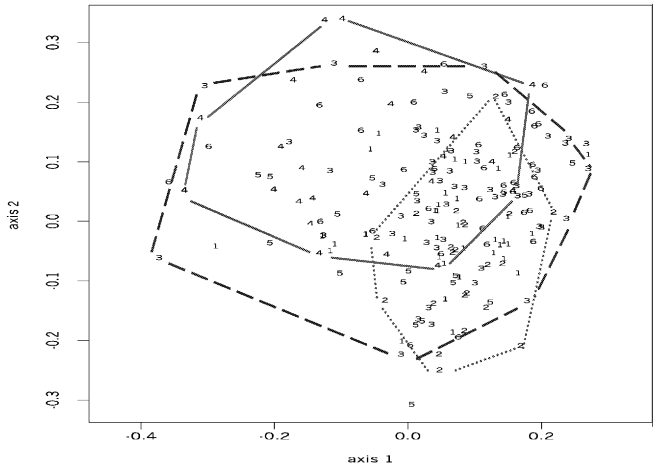

Fig. 7 Scatter plot of 193 accessions from six breeding companies: *1* Benoist, *2* Blondeau, *3* Desprez, *4* INRA, *5* Tourneur, *6* Vilmorin-Verneuil). (For the purpose of clarity, the *lines* (*convex hulls*) connect the extremes of the variation for only three breeders: – INRA, – - – - – - Desprez, ... Blondeau)

Variations between breeders

As for the temporal analysis, an AMOVA was applied to the six groups of breeders, to analyse genetic structuring between them. Although the among-breeder component is significant (Table 9), the results demonstrated that the most important variation is explained by the within-breeder component (95.6% of the total variation).

A study of allele number was carried out on these six breeder groups, using the rarefaction method. The total allele number ranged from 175 (Blondeau) to 225 (INRA). These results on allelic richness are in agreement with those on average gene diversity (data not shown); the breeder group which conserved the highest diversity level is INRA, while the lowest one appears to be Blondeau. Figure 6 separately summarises the differences in allele number between breeders for common and rare alleles. Common allele numbers ranged from 112 (Desprez) to 139 (Tourneur), while rare allele numbers varied from 53 (Blondeau) to 85 (INRA).

The results from non-metric multidimensional scaling of the distances between accessions of those six breeders are presented in Fig. 7. The first two axes account for 11.7% and 7.4% of the total variance, respectively. This figure provides a visual representation of the diversity present in each breeder group by drawing convex hulls around the extremes of each breeder. This approach highlights the ranges of diversity of the varieties within each breeding company. For reasons of clarity, only Blondeau, Desprez and INRA convex hulls are shown. It is clear that INRA and Blondeau varieties differ the most, while Desprez genetic material overlaps most of the diversity revealed in the six different breeder groups. The surface of the convex hull for Blondeau also indicated, once again, a lower differentiation within this group compared with the others.

Discussion

The results obtained from microsatellite analysis of these 559 French accessions of hexaploid wheat showed that microsatellites are very useful for determining genetic diversity. Our set of 41 microsatellite markers, detecting 42 polymorphic loci, produced a total number of 609 different alleles (without counting null alleles). This is the first time that such a large number of alleles are reported to describe genetic diversity in a wheat collection. Recently, Huang et al. (2002) reported a total number of 470 alleles, detected from 26 microsatellite loci, in 998 accessions of hexaploid bread wheat originating from 68 countries. Röder et al. (2002), studying 502 recent European wheat varieties, detected 199 alleles from 19 wheat microsatellites. In 2002, Zhang et al. (2002) studied 43 Chinese wheat varieties with 90 polymorphic SSR

markers and obtained 501 alleles. They tried to find an estimation of the minimum number of SSR alleles needed to reveal genetic relationships between these accessions. They found that the use of only 167 alleles made it possible for them to discriminate between all Chinese wheat genotypes, but that it was not enough to construct a stable dendrogram to objectively reflect genetic relationships. Their estimations suggested that 350–400 alleles were required to distinguish between all of their varieties of common wheat. On the other hand, other descriptions of cereal germplasm collections revealed that 249 alleles were enough to cluster 306 barley accessions (Baek et al. 2003), or that 163 alleles were sufficient to characterise 108 rice accessions (Gao et al. 2003). Therefore, our number of alleles (609) appears to be large enough to guarantee the robustness of our results.

The high correlation between gene diversity and the number of alleles (r=0.69, P<0.01) was previously observed by Huang et al. (2002). An average of 14.5 alleles per locus and a mean PIC value of 0.662 was detected in the present study; compared to Röder et al. (2002), the PIC value is equivalent but our average number of alleles is higher. This can probably be explained by the fact that we analysed both landraces and old varieties, not only recent ones.

The different contribution of three genomes to genetic variation was also confirmed in this study; surprisingly, the D genome presents a greater number of alleles than the B genome, while the greatest number was founded in the A genome. This is not in agreement with results from Huang et al. (2002) on 998 accessions, for which the B genome was more variable. In contrast, the same authors reported that the lowest genetic variation exists in the chromosomes of homologous group 4. A possible explanation is that many mutations occurring in these chromosomes might be lethal. They suggest that this could be due to one recessive gene and one dominant gene for male sterility, which are located on the 4BS and 4DS chromosomes of wheat, respectively.

Allelic diversity was observed for all accessions. Some loci display a regular normal distribution of allele frequencies (for instance *Xgwm11*, *Xgwm539*, *Xgwm120*, etc.), while other ones present plurimodal distributions (*Xgwm642*, *Xgwm161*, etc.).

According to Ohta and Kimura (1973) and to Di Rienzo et al. (1994), there are two mutational models for microsatellites: (1) the SMM implies that an allele mutates only by losing or gaining a single tandem repeat and can thus mutate towards an allele already present in the population; and (2) the IAM implies that a mutation always results in a new allele not yet encountered in the population, the size of which is independent of the initial size. In order to determine the level of diversity between groups, F_{st} values are currently estimated using the IAM, while R_{st} values are determined on the basis of the SMM.

It is generally considered that a normal distribution of allele frequency is related to an SMM, while irregular bimodal, trimodal or tetramodal distributions can be explained by an IAM. However, for some loci, we can also observe modifications in the distribution of their allelic frequencies which cannot be due to mutation phenomena. This is illustrated in Fig. 8, which represents the changes in allelic distribution over a period of time for *Xgwm642* loci: in landraces, allele '189' is a dominant allele. Then, in genetic material from the 1960s, the frequency of allele '203' increases. Finally, in varieties from the last decadal group, the frequencies of both alleles are quite similar. In that case, we can imagine that selection events over a period of time favour the '203' allele. In fact, as reported for other microsatellite loci, it appears that either direct or indirect selection can play a role in creating allelic variation (Korzun et al. 1999).

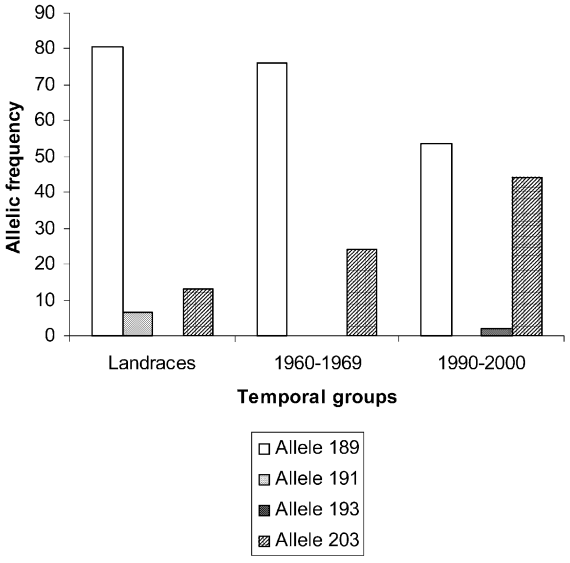

Fig. 8 Changes over time in allelic distribution for *Xgwm642* locus

Therefore, since it was difficult to assign a specific model to each microsatellite used in the present study, we decided to calculate the two different parameters, F_{st} and R_{st}, on the whole set of markers and to compare them. Results are very similar (r=0.98). The dendrogram clearly separates landraces from varieties registered after 1840. These varieties are clustered into two groups: one cluster includes accessions registered before 1970, while the last three decadal groups are in a second cluster. The two AMOVA tables supported these results; although most of the molecular diversity appeared within decadal groups of accessions, we observed significant effects for V_a and V_b components. However, in both cases, only a minor proportion of the overall variance appears to be the result of any temporal drift between decadal groups.

In order to look at variations in diversity related to temporal trends, allelic richness was calculated for each of the eight decadal groups. As demonstrated by Spencer et al. (2000), allelic diversity is one of the most sensitive parameters for looking at genetic changes that resulted from bottleneck effects. To standardise the results of allelic richness, we used the rarefaction technique (El Mousadik et al. 1996). We observed that this method is very effective for comparing allelic richness parameters

(*A*, *H*) between groups which have different sizes. As pointed out in the paper of Djé et al. (2000), without this method, no comparison could be possible because of sample size effect.

Here, the decrease of allelic richness between landraces and all the decadal groups appears to be significant. Similar observations have been reported on sunflower (Tang et al. 2003) where the authors observed greater allelic diversity in landraces and wild populations compared with domesticated and selected material. For decadal groups, allelic richness is always significantly lower during the 1960–1970 period, compared with the other decadal groups. Thus, these results confirm the diversity structure observed on the dendrogram for F_{st}, with a clear clustering, on the one hand, between landraces and varieties and, on the other, between the before-1970s and after-1970s decadal groups.

This threshold of 1970 can be probably linked to the 'green revolution' period: after the Second World War, European countries needed to increase food production and also realised the importance of creating varieties which could be produced in less developed countries. Cereal crop breeding programmes led them to create new and much more productive varieties able to valorise high level of inputs. Those 'intensive' varieties were probably bred during the 1960s, using a reduced number of progenies chosen among the previous registered varieties. This produced a sort of bottleneck effect due to the extensive use of a few genitors which provided new useful characteristics such as reduced height to facilitate increased nitrogen fertilisation (*rht* genes). This could be one of the reasons why we observed a decrease in genetic diversity during the 1960s. Finally, at the end of the twentieth century, allelic richness increased slowly, compared to the 1960s. This could be explained by the use of new foreign progenies and the development of genetic resource exchanges between an increasing number of breeders during this period of time. As for landraces, they were at the origin of the first plant-breeding programmes during the nineteenth century, and it seems quite logical to observe a decrease in genetic diversity between source material and varieties in a plant-breeding programme, due to strong selection pressure.

To determine qualitative variations in allelic richness, rare allele composition was also observed in each group. Once again, the greatest diversity appears in landrace groups. We therefore decided to first compare landraces with the other decadal groups (Fig. 4).

The curve (highest line) of the accumulation of different alleles shows that there is an overall increase in the differences between landraces and other groups over a period of time. This indicates that the global composition of rare alleles is always changing when compared with landraces. Recent varieties differ to an even greater extent from the landraces. This could probably be explained by the recent introduction of new genes (for instance: dwarf genes, resistance genes, etc.) and the intensive use of related species or foreign cultivars during the last decades. The goals of breeding programmes have also changed since the middle of the nineteenth century, and new alleles have probably been selected.

However, as indicated by Fig. 5, the evolution of rare allele composition was not the same during the different periods. With the exception of a short period corresponding to the Second World War, we can clearly observe an important change during the 1960s: globally, rare alleles appearing in French varieties are greater before the 1960s and, conversely, the number of disappearing alleles is at its highest after that period. Once again, this change could be related to the 'green revolution' period during which intensive selection pressure led to the loss of qualitative diversity. However, if we look at the last results between 1990 and 2000, we can see a slight increase in the appearance of new alleles. This is perhaps the indication of a new diversification period which could be explained by the use of new exotic germplasm in plant-breeding programmes.

All of the results concerning temporal trends in the diversity of cultivated crops are quite consistent with those from Donini et al. (2000), who observed similar effects when comparing 55 UK wheat accessions registered over the period 1934–1995. They are also in agreement with the results of Koebner et al. (2003) concerning a collection of 134 UK barley varieties over the period 1925–1995. In both cases, it has been consistently demonstrated that plant breeding has resulted in a qualitative, rather than a quantitative, shift in the diversity of cereal crops over time in the United Kingdom. In our case, considering only French registered varieties, changes in diversity related to temporal flux also appear to be more qualitative than quantitative, except for a certain narrowing at various times, particularly at the end of the 1960s.

By contrast, when we compared French varieties of a given period with landraces, we observed a global decrease of allelic richness (about 25%) between landraces and selected varieties. Even if microsatellite diversity is considered to be neutral, a selection effect may have occurred, due to the proximity of microsatellites with selected genes. We can also imagine, as suggested by Allard (1996), that this reduction in allelic diversity was largely due to the elimination of closed deleterious genes, rather than the erosion of useful genetic variations. Such a reduction was also observed by Russell et al. (2000), when comparing the most recent cultivars of European barley with a group of 19 landraces, considered as being founder genotypes. They concluded that there was a reduction in the spectrum of alleles at 28 SSR loci over a period of time, and a decrease in the level of diversity between landraces and modern varieties.

In order to observe a possible breeding-origin effect on the genetic structure of diversity within our collection of French bread wheat accessions, we limited ourselves to the six breeder companies which have produced a large number of varieties over a relatively long period. In this way, we avoided confusion between temporal and breeding-origin effect. Unfortunately, although the between-

breeders component of AMOVA is significant, most of the diversity is within breeder groups. The results on allelic richness give us more details concerning a possible qualitative variation in allelic composition of the different breeders. INRA genetic material seems to be the most diversified, while Blondeau clearly presents a lower level of diversity. This may be explained by differences in the aims and policies of the two breeders: INRA, which is a public organisation, is more involved in conservation, diversification and the use of wide-ranging genetic resource collections as opposed to Blondeau, which is, as a private company, mainly interested in creating new cultivars and using intensive selection pressure for specific targets.

However, differences in breeder programmes do not appear to be very significant in this study. These similarities may be explained by the fact that a lot of cultivars are probably related. Unfortunately, pedigrees are often unknown for recent varieties and kinship coefficients cannot be estimated.

Conclusion

The present study of French bread wheat genetic diversity shows that microsatellite markers are a very effective tool for evaluating large collections of genetic resources. French accessions appear to be considerably polymorphic and to possess many rare alleles which may be very useful for broadening diversity and for the improvement of new cultivars.

The approaches used in this work also provide important information concerning the effect of modern plant breeding and agriculture practices on the diversity of cereal crops. After consideration of a wide collection of cultivars over two centuries of cultivation, we observed some evidence of changes in diversity related to temporal variations. When compared with old landraces, registered cultivars present a decrease of about 25% in their allelic richness. However, when comparisons are made between decadal groups of varieties, changes in diversity appear more qualitative than quantitative, except at the end of the 1960s, when a bottleneck effect may have occurred. Differences between landraces are more and more important over time, but new varieties appear to be increasingly similar to each other in relation to their allelic composition. Finally, when varieties are compared according to their breeding origin, few differences are observed between global genetic materials of these French breeding companies.

These results lead to several conclusions. The first one concerns plant breeders who should increase the exchange and use of genetic resources for broadening genetic material and improving new cultivars. Otherwise, the present evolution could be prejudicial to the long-term maintenance of genetic diversity. Another consequence is related to the conservation of crop genetic resources. For instance, in the case of the constitution of a bread wheat core collection, it would be necessary to sample accessions not only according to their geographical or breeding origins, but also according to their registration period in particular. And probably, if we used an M strategy (Schoen et al. 1993) which consists of maximising allelic diversity, this would lead us to sample more accessions in landraces than in cultivars, since allelic richness is more pronounced in landraces. These approaches are under investigation and will be published in further papers.

Acknowledgements The authors would like to thank Florence Exbrayat-Vinson and Nathalie Marty for their technical assistance. This research was supported by the French Ministry of Research and Technologies and the Ministry of Finances (ASG programme: Cereal Genotyping and Food Quality).

References

Allard RW (1996) Genetic basis of the evolution of adaptedness in plants. Euphytica 92:1–11

Baek HJ, Beharav A, Nevo E (2003) Ecological-genomic diversity of microsatellites in wild barley, *Hordeum spontaneum*, populations in Jordan. Theor Appl Genet 106:397–410

Ben Amer IM, Borner A, Röder MS (2001) Detection of genetic diversity in Lybian wheat genotypes using wheat microsatellite markers. Genet Resour Crop Evol 48:579–585

Branlard G, Dardevet M, Saccomano R, Lagoutte F, Gourdon J (2001) Genetic diversity of wheat storage proteins and bread wheat quality. Euphytica 119:59–67

Bustarret J (1944) Variétés et variations. Ann Agr XIV:336–362

Cao W, Scoles G, Hucl P, Chibbar RN (2000) Phylogenetic relationships of five morphological groups of hexaploid wheat (*Triticum aestivum* L. em Thell.) based on RAPD analysis. Genome 43:724–727

Chandrashekhar P, Joshi P, Nguyen HT (1993) Application of the random amplified polymorphic DNA technique for the detection of polymorphism among wild and cultivated tetraploid wheats. Genome 36:602–609

Christiansen MJ, Andersen SB, Ortiz R (2002) Diversity changes in an intensively bread wheat germplasm during the 20th century. Mol Breed 9:1–11

Dagnélie P (1986) Théorie et méthodes statistiques, vol 2. Duculot, Gembloux, France, p 385

DeLacy IH, Skovmand B, Huerta J (2000) Characterization of Mexican wheat landraces using agronomically useful attributes. Genet Resour Crop Evol 47:591–602

Dillmann C, Bar-Hen A, Guérin D, Charcosset A, Murigneux A (1997) Comparison of RFLP and morphological distances between maize *Zea mays* L. inbred lines. Consequences for germplasm protection purposes. Theor Appl Genet 95:92–102

Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB (1994) Mutational processes of simple-sequence repeat loci in human populations. Proc Natl Acad Sci USA 91:3166–3170

Djé Y, Heuertz M, Lefebvre C, Vekemans X (2000) Assessment of genetic diversity within and among germplasm accessions in cultivated sorghum using microsatellite markers. Theor Appl Genet 100:918–925

Donini P, Law JR, Koebner RM, Reeves JC, Cooke RJ (2000) Temporal trends in the diversity of UK wheat. Theor Appl Genet 100:912–917

El Mousadik A and Petit RJ (1996) High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. Theor Appl Genet 92:832–839

Gao L, Schaal BA, Jia J, Dong Y (2003) Assessment of population genetic structure in common wild rice *Oryza rufipogon* Griff. using microsatellite and allozyme markers . Theor Appl Genet 106:173–180

Guadagnuolo R, Savova Bianchi D, Felber F (2001) Specific genetic markers for wheat, spelt, and four wild relatives: comparison of isozymes, RAPDs, and wheat microsatellites. Genome 44:610–621

Guyomarc'h H, Sourdille P, Charmet G, Edwards KJ, Bernard M (2003) Characterization of polymorphic microsatellite markers from *Aegilops tauschii* and transferability to the D-genome of bread wheat. Theor Appl Genet 104:1164–1172

Huang XQ, Borner A, Röder MS, Ganal MW (2002) Assessing genetic diversity of wheat (*Triticum aestivum* L.) germplasm using microsatellite markers. Theor Appl Genet 105:699–707

Jones CJ, Edwards KJ, Castaglione S, Winfield MO, Sala F, van de Wiel C, Bredemeijer G, Vosman B, Matthes M, Daly A, Brettschneider R, Bettini P, Buiatti M, Maestri E, Malcevschi A, Marmiroli N, Aert R, Volckaert G, Rueda J, Linacero R, Vazquez A, Karp A (1997) Reproductibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. Mol Breed 3:381–390

Kim HS, Ward RW (1997) Genetic diversity in Eastern U.S. soft winter wheat (*Triticum aestivum* L. em. Thell.) based on RFLPs and coefficients of parentage. Theor Appl Genet 94:472–479

Koebner RM, Donini P, Reeves JC, Cooke RJ, Law JR (2003) Temporal flux in the morphological and molecular diversity of UK barley. Theor Appl Genet 106:550–558

Korzun V, Röder MS, Wendehake K, Pasqualone A, Lotti C, Ganal MW, Blanco A (1999) Integration of dinucleotide microsatellites from hexaploid bread wheat into a genetic linkage map of durum wheat. Theor Appl Genet 98:1202–1207

Kruskal JB, Wish M (1978) Multidimensional scaling. Sage, Beverly Hills, Calif.

Li Z, Zhang H, Zeng Y, Yang Z, Shen S, Sun C, Wang X (2002) Studies on sampling schemes for the establishment of core collection of rice landraces in Yunnan, China. Genet Res Crop Evol 49:67–74

Liu F, Sun GL, Salomon B, von Bothmer R (2002) Characterization of genetic diversity in core collection accessions of wild barley, *Hordeum vulgare* ssp. *spontaneum*. Hereditas 136:67–73

Lund B, Ortiz R, Skovgaard IM, Waugh RW, Anderson SB (2003) Analysis of potential duplicates in barley gene bank collections using re-sampling of microsatellite data. Theor Appl Genet 106:1129–1138

Manifesto MM, Schlatter AR, Hopp HE, Suarez HE, Dubcovsky J (2001) Quantitative evaluation of genetic diversity in wheat germplasm using molecular markers. Crop Sci 41:682–690

Matus IA, Hayes PM (2002) Genetic diversity in three groups of barley germplasm assessed by simple sequence repeats. Genome 45:1095–1106

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci USA 76:5269–5273

Ni J, Colowit PM, Mackill DJ (2002) Evaluation of genetic diversity in rice subspecies using microsatellite markers. Crop Sci 42:601–607

Ohta T, Kimura M (1973) A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. Genet Res 22:201–204

Petit RJ, El Mousadik A, Pons O (1998) Identifying populations for conservation on the basis of genetic markers. Cons Biol 12:844–855

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

Prasad M, Varshney RK, Roy JK, Balyan HS, Gupta PK (2000) The use of microsatellites for detecting DNA polymorphism, genotype identification and genetic diversity in wheat. Theor Appl Genet 100:584–592

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

Röder MS, Wendehake K, Korzun V, Bredemeijer G, Laborie D, Bertrand L, Isaac P, Rendell S, Jackson J, Cooke RJ, Vosman B, Ganal MW (2002) Construction and analysis of a microsatellite-based database of European wheat varieties. Theor Appl Genet 106:67–73

Russel JR, Ellis RP, Thomas WTB, Waugh R, Provan J, Booth A, Fuller J, Lawrence P, Young G, Powell W (2000) A retrospective analysis of spring barley germplasm development from 'foundation genotypes' to currently successful cultivars. Mol Breed 6:553–568

Schneider S, Kneffer JM, Roessle D, Excoffier L (1996) Arlequin. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, Switzerland

Schoen DJ, Brown AH (1993) Conservation of allelic richness in wild crop relatives is aided by assessment of genetic markers. Proc Natl Acad Sci USA 90:10623–10627

Sjakste TG, Rashal I, Röder MS (2003) Inheritance of microsatellite alleles in pedigrees of Latvian barley varieties and related European ancestors. Theor Appl Genet 106:539–549

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457–62

Spencer CC, Neigel JE, Leberg PL (2000) Experimental evaluation of the usefulness of microsatellite DNA for detecting demographic bottlenecks. Mol Ecol 9:1517–1528

Stachel M, Lelley T, Grausgruber H, Vollman J (2000) Application of microsatellites in wheat (Triticum aestivum L.) for studying genetic differentiation caused by selection. Theor Appl Genet 100:242–248

Tang S, Knapp SJ (2003) Microsatellites uncover extraordinary diversity in Native American land races and wild populations of cultivated sunflower. Theor Appl Genet 106:990–1003

Ward RW, Yang ZL, Kim HS, Yen C (1998) Comparative analysis of RFLP diversity in landraces of *Triticum aestivum* and collections of *T. tauschii* from China and Southwest Asia. Theor Appl Genet 96:312–318

Zhang XY, Li CW, Wang LF, Wang HM, You GX, Dong YS (2002) 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. Theor Appl Genet 106:112–117