Theor Appl Genet (2002) 465–472

J.K. Roy · H.S. Balyan · M. Prasad · P.K. Gupta

Use of SAMPL for a study of DNA polymorphism, genetic diversity and possible gene tagging in bread wheat

Received: 5 April 2001 / Accepted: 17 May 2001

Abstract **S**elective **A**mplification of **M**icrosatellite **P**olymorphic Loci (SAMPL) technology was used in bread wheat for the first time for a study of genetic diversity, genotype identification and gene tagging. The diversity studies involved 55 wheat genotypes and two SAMPL primer pairs (SAMPL-6 and SAMPL-7, each with a M-CAG primer), which together gave 43 polymorphic bands out of a total of 87 SAMPL bands. The average polymorphic information content (PIC) of SAMPL primers was 0.221 and that of SAMPL markers was 0.264. The marker index of SAMPL markers was 9.61. The genetic similarity (GS) coefficients for 1,485 pairs of genotypes ranged from 0.35 to 0.96 with an average of 0.65. A dendrogram was prepared on the basis of a similarity matrix using the UPGMA algorithm, which corresponded well with the results of principal component analysis (PCA). From a total of 55 genotypes, 54 could be distinguished using the SAMPL banding patterns of both primers. For gene tagging, 568 bands from a total of 1,185 SAMPL bands detected polymorphism between each of the three pairs of parents differing for grain protein content (GPC), pre-harvest sprouting tolerance (PHST) and grain weight (GW). An association of six bands with GPC, of seven bands with PHST and four bands with GW was observed using bulked segregant analysis (BSA).

Keywords SAMPL · DNA polymorphism · Diversity · PIC · Jaccard similarity coefficient · Principal component analysis

Communicated by H.F. Linskens

J.K. Roy · H.S. Balyan · M. Prasad · P.K. Gupta (✉)
Molecular Biology Laboratory,
Department of Agricultural Botany, Ch. Charan Singh University,
Meerut – 250 004, U.P., India
e-mail: pkgupta@ndf.vsnl.net.in
Tel.: 0091-121-768195, 121-762505
Fax: 0091-121-760577, 121-767018

Present address: M. Prasad, Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK), Correnstraße 3, D-06466 Gatersleben, Germany

Introduction

During the last two decades, DNA markers have been extensively used in all major crops for detecting DNA polymorphisms leading to the preparation of genetic maps, the tagging of genes for economic traits and the estimation of genetic diversity. Bread wheat, due to its large genome size (16×10^9 bp) and a very high proportion of repetitive DNA (>80%), was initially considered to be difficult material for such studies. However, significant progress was later made in this crop, using the two first-generation molecular marker systems, which included restriction fragment length polymorphisms (RFLPs) and random amplified polymorphic DNAs (RAPDs). The limitations of RFLPs and RAPDs were, however, soon recognized, and two second-generation PCR-based marker systems became popular. These included the simple sequence repeats (SSRs) and the amplified fragment length polymorphisms (AFLPs) (Roder et al. 1998; Bohn et al. 1999; Prasad et al. 1999, 2000; Roy et al. 1999; Varshney et al. 2000; for reviews see Gupta et al. 1999; Gupta and Varshney 2000). Although SSRs have the advantage of being locus-specific and multi-allelic, they need cloning and sequencing for designing the primers (Weissenbach et al. 1992; Morgante and Olivieri 1993; Powell et al. 1996). The AFLP approach, on the other hand, needs no cloning and sequencing, and gives a very large number of scorable fragments, which enhance its power to detect polymorphism. However, AFLP bands are sometimes too many to be conveniently scored and the markers are often dominant, thus making the choice between SSRs and AFLPs rather difficult. Therefore, some of the merits of SSR and AFLP markers have been combined into a single assay, called selective amplification of microsatellite polymorphic loci or SAMPL (Morgante and Vogel 1994). SAMPL is a modified AFLP technique in which adapter-ligated restriction fragments of a conventional AFLP are used, but the final PCR amplification is achieved using primers which differ from those employed in AFLP. One of the primers is a 5′ self-anchoring compound SSR (abundant in many

plant genomes), and the other is an AFLP primer that is designed on the basis of the sequences of the synthetic adapter plus the restriction site, and carries 2–3 selective nucleotides (Witsenboer et al. 1997). This technique has a high multiplex ratio like AFLP, and has some degree of locus specificity like SSRs, thus making it the most efficient of all the molecular markers known so far (except of course SNPs, that have yet to be used in plant systems in any significant measure). However, SAMPL has been sparingly used in the past; the only examples of plant systems where it was successfully utilized being lettuce, carrot and a conifer (Witsenober et al. 1997; Paglia and Morgante 1998; Vivek and Simon 1999). In this communication, we report the use of SAMPL in bread wheat for the first time, demonstrating its utility for studies involving genetic diversity, genotype identification and gene tagging.

Materials and methods

Seed material

The material used for the diversity study included seeds of 55 wheat genotypes, that originated in 29 countries belonging to six continents. The seed material of these genotypes was procured from the Directorate of Wheat Research, Karnal, India. The details of the genotypes with the countries of origin are given in Fig. 2, and their pedigrees (as far as known) are described in Prasad et al. (1999).

For the tagging of genes/QTLs for three different grain quality traits, including grain protein content (GPC), pre-harvest sprouting tolerance (PHST) and grain weight (GW), the seed material consisted of the following three pairs of bread wheat genotypes, each differing for a different quality trait, and each used for deriving a separate set of recombinant inbred lines (RILs). For GPC, the two parents were PH132 (high) and WL711 (low), for PHST they were HD2329 (susceptible) and SPR8198 (tolerant), and for GW they were Rye Selection 111 (high) and Chinese Spring (low). Each of the three sets of 100 RILs for GPC, PHST and GW were derived following the single-seed descent method, at Punjab Agricultural University (PAU), Ludhiana.

DNA extraction

DNA was extracted from leaves of 1-month-old field-grown single plants following a modified CTAB method (Saghai Maroof et al. 1984). The concentration of isolated DNA was determined on agarose gels using known concentrations of lambda DNA.

AFLP adapters and AFLP/SAMPL primers

The AFLP analysis system I (kit) from Life Technology, USA, was used, which included the adapters, pre-amplification primers and selective amplification primers. The two SAMPL primers were synthesized, on contract, by Oswel, University of Southampton, UK. The details of the adapter sequences and the different primers are given in Table 1.

SAMPL analysis

SAMPL analyses were performed following the AFLP technology suggested by Vos et al. (1995) with some modifications. SAMPL primers, labelled with γ-^{32}p-ATP, were used in place of one of the two AFLP primers. Genomic DNA of each genotype was digested with the restriction enzymes *Eco*RI and *Mse*I. All PCR reactions were performed in a Perkin Elmer thermal cycler following Vos et al. (1995). Each PCR amplified product was mixed with an equal volume of formamide loading dye, denatured for 5 min in a boiling water bath and then run in a 6% polyacylamide gel at constant voltage (1,200 V) until the slower dye was two-thirds down the length of the gel. The gel was transferred to 3MM Whatman chromatography paper and dried on a gel drier for 2 h at 80°C. The dried gel was used for exposing an X-ray-film overnight (or longer as required) at –80°C.

Bulked segregant analysis (BSA)

BSA was conducted following Michelmore et al. (1991). The two bulks for each trait were prepared by pooling the DNA from 5 to 8 RILs representing each of the two tails of the normal distribution, in case of GPC as well as GW, and from five RILs each randomly chosen from tolerant and susceptible groups of RILs, in case of PHST.

Table 1 Sequences of adapters and primers

Adapter/primer	Code	Sequence
Adapters		
*Eco*RI adapter		5′-CTCGTAGACTGCGTACC CATCTGACGCATGGTTAA-5′
*Mse*I adapter		5′-GACGATGAGTCCTGAG TACTCAGGACTCAT-5′
Preamplification primers		
*Eco*RI primer+1	E_A	5′-GACTGCGTACCAATTC**A**-3′
*Mse*I primer+1	M_C	5′-GATGAGTCCTGAGTAA**C**-3′
Selective amplification primers		
*Mse*I primer+3	M-CAG	5′-GATGAGTCCTGAGTAA**CAG**-3′
*Mse*I primer+3	M-CTT	5′-GATGAGTCCTGAGTAA**CTT**-3′
*Mse*I primer+3	M-CTC	5′-GATGAGTCCTGAGTAA**CTC**-3′
*Mse*I primer+3	M-CAT	5′-GATGAGTCCTGAGTAA**CAT**-3′
*Mse*I primer+3	M-CAA	5′-GATGAGTCCTGAGTAA**CAA**-3′
SAMPL primers		
SAMPL-6		5′-ACACACACACACACATATAA-3′
SAMPL-7		5′-TGTGTGTGTGTGTGTATAT-3′

Data analysis

*Polymorphic information content (*PIC*)*

The probability of detection of polymorphism by a marker system between two randomly drawn genotypes used in a study is described as PIC or average heterozygosity (H_{av}), although for self-pollinated crops like bread wheat, the term average heterozygosity appears to be a misnomer. PIC was calculated separately for each SAMPL primer and for each SAMPL band. *PIC* or H_{av} for each primer was obtained using the following formula and averaged:
H_{av} (*PIC*)=$\beta\sum H_p/n_p$
, where H_{av} is average heterozygosity (*PIC*) for all SAMPL bands (both polymorphic and monomorphic), β is the proportion of polymorphic bands, H_p=1–$\sum p_i^2$ (p_i=proportion of genotypes having a i^{th} band), n_p=number of polymorphic bands. In the above formula, $\sum H_p/n_p$ is the average heterozygosity for only the polymorphic bands, and needs to be multiplied with β to obtain the desired value of *PIC*. H_{av} *(PIC)* is thus the probability of any solitary band being polymorphic, when examined in the given sample of genotypes (for more details see Powell et al. 1996).

The *PIC* of each marker band *(i)*, which is the probability of any band being polymorphic between two genotypes drawn randomly, was calculated using the following formula (Roldan-Ruiz et al. 2000): $PIC_i=2f_i(1–f_i)$, where f_i is the frequency of genotypes showing the presence of band *i.*

*Marker index (*MI*)*

The *MI* was calculated using the formula, *MI*=Average *PIC* of SAMPL primers (H_{av}, as above)×average number of loci per assay unit.

Genetic similarity (GS)

For the purpose of assessing genetic diversity leading to cluster analysis, the data in binary format (1, 0) were used to compute pair-wise similarity coefficients (Jaccard 1908), utilizing the SIMQUAL (similarity for qualitative data) method in NTSYS-pc software.

Cluster analyses

A similarity coefficient matrix was used for cluster analysis following UPGMA (unweighted pair group method with arithmetic averages), which is one of the several SAHN (sequential, agglomerative, hierarchical, and nested) clustering methods that are available (Sneath and Sokal 1973). NTSYS-pc software was used for analysis and the resulting clusters were represented in the form of a dendrogram.

Principal component analysis (PCA)

The original 1–0 data matrix was used for calculating a correlation matrix between pairs of markers. The correlation matrix was employed for the calculation of eigen values, which were then used for determining the coordinates for each genotype that were used for PCA.

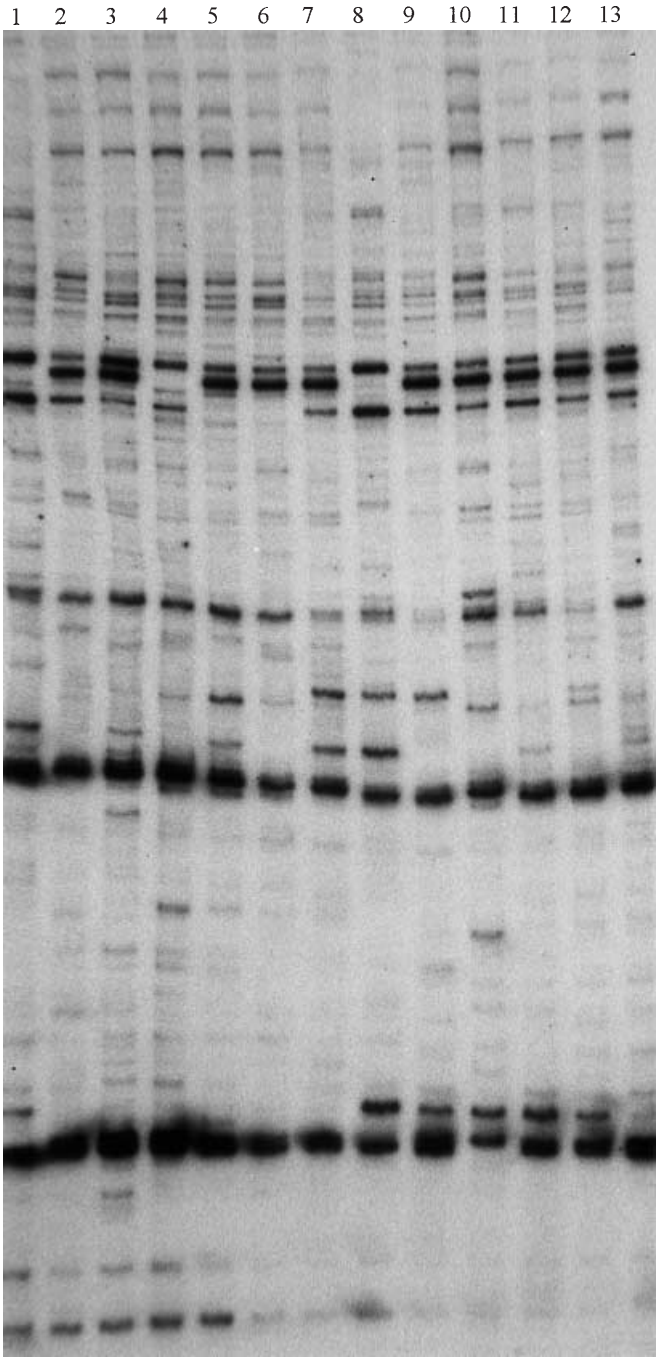

Fig. 1 Representative SAMPL profiles, obtained using SAMPL-7 and M-CAG primers, for 1–13 bread wheat genotypes, serially representing E271, E549, E319, E336, E581, E585, E661, E677, E680, E780, E965, E1000 and E1003

Results

DNA polymorphism

Two SAMPL primers (SAMPL-6 and SAMPL-7), each in combination with the AFLP primer M-CAG, were tried with 55 elite wheat genotypes for a study of genetic diversity. SAMPL-6 gave a total of 42 bands, containing 22 polymorphic bands, and SAMPL-7 gave a total of 45 bands, containing 21 polymorphic bands, thus giving a total of 43 polymorphic bands out of a total of 87 bands (Fig. 1).

A total of five primer pairs, each using only one of the two above SAMPL primers (SAMPL-6 with five *Mse*I primers; see Table 2) were also used to study poly-

Table 2 The details of SAMPL primers used, the frequencies of bands scored and the putative markers identified during the study of polymorphism between parents differing for GPC, PHST and GW. The total number of bands is given under GPC, PHST and GW, and the percentage of polymorphism is given in parentheses

SAMPL primers	GPC	Putative markers[a]	PHST	Putative markers[a]	GW	Putative markers[a]
SAMPL-6×M-CAG	74 (43.24)	*XccuES$_6$400* *XccuES$_6$390* *XccuES$_6$300*	74 (37.83)	*XccuES$_6$430* *XccuES$_6$370*	60 (46.67)	–
SAMPL-6×M-CTT	112 (51.78)	*XccuES$_6$350*	83 (49.39)	*XccuES$_6$410* *XccuES$_6$345* *XccuES$_6$280*	80 (52.50)	*XccuES$_6$355* *XccuES$_6$340* *XccuES$_6$330*
SAMPL-6×M-CTC	129 (43.41)	*XccuES$_6$365* *XccuES$_6$325*	92 (39.13)	–	80 (51.25)	–
SAMPL-6×M-CAT	94 (51.06)	–	87 (43.67)	*XccuES$_6$150* *XccuES$_6$110*	73 (54.79)	*XccuES$_6$320*
SAMPL-6×M-CAA	57 (49.12)	–	40 (62.50)	–	50 (54.00)	–
Total	466 (47.63)	6	376 (44.68)	7	343 (51.89)	4

[a] Putative markers are designated as far as possible on the basis of nomenclature for AFLP proposed at 9th Int Wheat Genet Symp; X is a symbol for the DNA marker, ccu for CCS University; E for the rare cutter enzyme *Eco*RI used, S_6 is a SAMPL primer and the three digits at the end represent the mol wt in base pairs

morphism between three pairs of parent genotypes differing for GPC, PHST and GW. A large number of bands ranging from 50 bp to 450 bp were available exhibiting sufficient polymorphism within each pair of parents. However, only clear and distinct bands were scored for further data analysis. Between the two parents differing for each of the three traits, 222 of 466 bands were polymorphic for GPC, 168 of 376 bands were polymorphic for PHST, and 178 of 343 bands were polymorphic for GW (Table 2).

Genetic diversity and genotype identification

For the two SAMPL primers employed with 55 bread wheat genotypes, the average PIC was 0.221, and the MI was 9.61. The PIC values of individual SAMPL bands ranged from 0.035 to 0.499 with an average of 0.264. Genetic similarity (GS) coefficients for 1,485 possible pairs of genotypes, based on the above 43 polymorphic bands, ranged from 0.35 in a solitary pair of genotypes (E288 and E3389) to 0.96 in each of two pairs (between E4229 and E4813, and between E3839 and E3901), giving an average GS value of 0.65. A genetic similarity matrix prepared on the basis of above GS values was used for cluster analysis through UPGMA resulting in a dendrogram (Fig. 2). The genotypes were grouped into two clusters, cluster I containing 54 genotypes and cluster II containing a solitary genotype, i.e. E3840 from Columbia. The 54 genotypes belonging to cluster I were again grouped into two sub-clusters, sub-cluster Ia containing 53 genotypes and Ib containing another solitary genotype, E2055, from Poland.

The results of principal component analysis are shown in Fig. 3. Of the total polymorphism, only 21.51% was accounted for by the first two components.

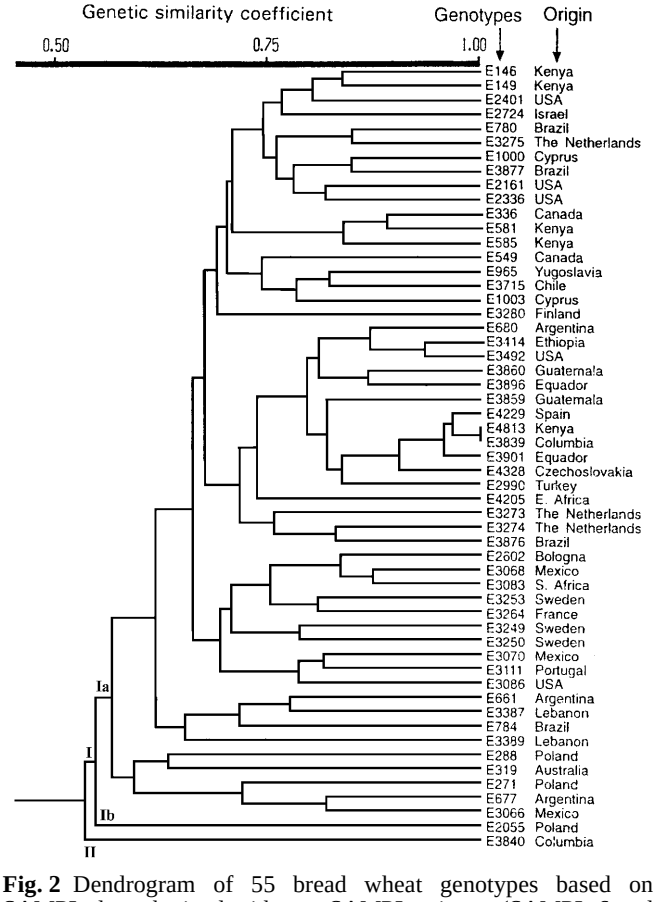

Fig. 2 Dendrogram of 55 bread wheat genotypes based on SAMPL data obtained with two SAMPL primers (SAMPL-6 and SAMPL-7)

The most-divergent genotypes, namely E3840 from Columbia and E2055 from Poland, were well separated from the other genotypes. The remaining 53 genotypes were clustered into three groups. The results of PCA thus

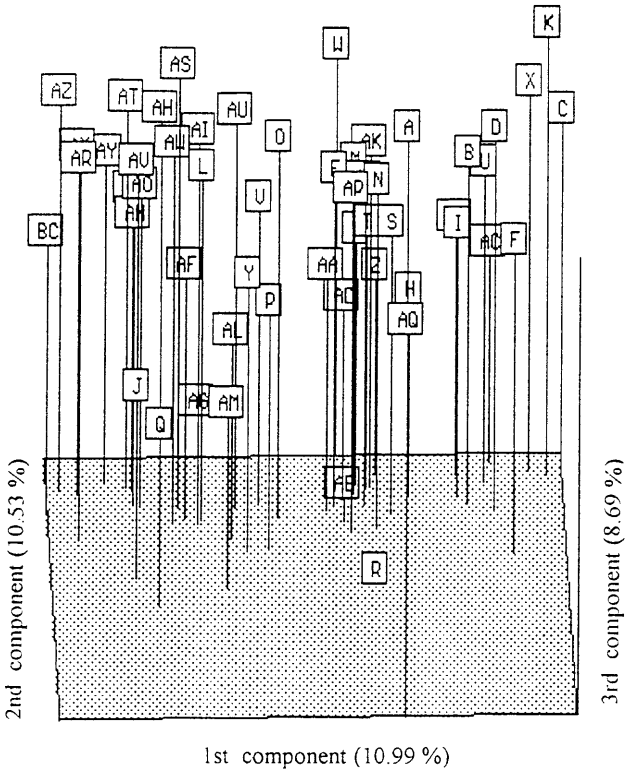

Fig. 3 Patterns of relationships among 55 bread wheat genotypes revealed by principal component analysis based on SAMPL data. The genotypes from A to BC are A-E146, B-E149, C-E288, D-E271, E-E549, F-E319, G-E336, H-E581, I-E585, J-E661, K-E677, L-E680, M-E780, N-E965, O-E1000, P-E1003, Q-E784, R-E2055, S-E2161, T-E2336, U-E2401, V-E2602, W-E2724, X-E3066, Y-E3068, Z-E3070, AA-E3083, AB-E3086, AC-E3249, AD-E3111, AE-E3250, AF-E3253, AG-E3264, AH-E3273, AI-E3274, AJ-E3275, AK-E3280, AL-E3387, AM-E3389, AN-E3414, AO-E3492, AP-E3715, AQ-E3840, AR-E3859, AS-E3860, AT-E3876, AU-E3877, AV-E3896, AW-E4205, AX-E4229, AY-E4813, AZ-E2990, BA-E3839 BB-E3901 and BC-E4328. Note: due to overlapping, the labels of four genotypes (G-E336, AE-E3250, BA-E3839 and BB-E3901) can not be clearly visualised in the figure

largely corresponded to those from cluster analysis obtained through UPGMA.

Unique patterns were obtained in 47 of the 55 genotypes using SAMPL-6, and in 49 of the 55 genotypes using SAMPL-7. When the results of both SAMPL primers are considered together, 54 of the 55 genotypes were found to have unique patterns in at least one of the two primers and 42 genotypes had unique patterns with each of the two primers. Only one genotype (E3414 from Ethiopia) of the 55 genotypes used did not have a unique pattern with either of the two primers. These results suggest that SAMPL technology can be effectively utilized for distinguishing a fairly large number of genotypes.

Bulked segregant analysis (BSA)

The five primer pairs that were used for detecting polymorphism between the parents differing for GPC, PHST and GW, as above, were also used for conducting BSA (Michelmore et al. 1991) on corresponding bulked DNA samples each for GPC, PHST and GW. With these primer pairs, apparent association of six SAMPL markers (bands) with GPC, seven with PHST and four with GW was observed (Figs. 4a, b and c). These bands showing associations with GPC, PHST and GW belonged to four of the five SAMPL primer pairs used in the present study (Table 2) and, of these four, only one (SAMPL-6 with M-CTT) gave bands which included at least one band showing association with each of the above three traits.

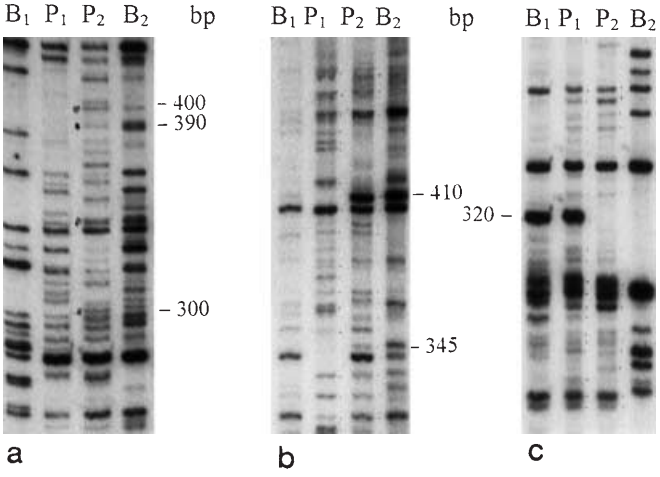

Fig. 4 Representative bulk segregant analysis (BSA) for GPC, PHST and GW using SAMPL primer-6 and AFLP primers M-CAG (**a**), M-CTT (**b**) and M-CAT (**c**). In each case, P_1 and P_2 are parents and B_1 and B_2 are corresponding bulks. (**a**) GPC: P_1=PH132 (high GPC) and P_2=WL711 (low GPC); (**b**) PHST: P_1=SPR8198 (tolerant) and P_2=HD2329 (susceptible); (c) GW: P_1=Rye Selection (high GW) and P_2=Chinese Spring (low GW)

Discussion

In the present study, the SAMPL approach has been used in bread wheat for the first time for a study of genetic diversity and gene tagging. Two SAMPL primer combinations (SAMPL-6 and SAMPL-7, each with the M-CAG AFLP primer; see Table 1) were found to be adequate for the study of genetic diversity among 55 different wheat genotypes, although for gene tagging five primer combinations were tried (Table 2). A higher number of primers for gene tagging was used, since screening of a much larger number of bands was required to identify at least a few bands that may be associated with each of the three traits of interest (Table 2). In an earlier comparable study involving DNA fingerprinting in a conifer (Norway spruce, *Picea abies*), as well as in the present study, two SAMPL primers (SAMPL-7 was used in both the studies) were found to be adequate for DNA fingerprinting (Paglia and Morgante 1998). Despite the similarity between bread wheat and Norway spruce both in genome size ($15–25\times10^9$ bp) and a high proportion of repetitive DNA (>80%), the level of polymorphism in the present

study (43 polymorphic bands) was almost double that observed earlier in Norway spruce (20 polymorphic bands). This was true even when the same primer, SAMPL-7, was employed in both studies (21 polymorphic bands in the present study as against 11 polymorphic bands in the earlier study), suggesting that the level of polymorphism is not entirely dependent on the SAMPL primer, but also depended on the plant system being examined. The comparison thus also suggested that SAMPL may prove to be relatively more useful for the analysis of the wheat genome than for the analysis of other large genomes represented by some conifers. An analysis of earlier SAMPL studies also suggested that the same SAMPL primer may give different levels of polymorphism in different plant systems. For instance, SAMPL primers with an AT extension (e.g. SAMPL-7) detected the highest level of polymorphism in corn and soybean, but not in lettuce (see Witsenboer et al. 1997).

When SAMPL technology is compared with SSR, one should also realize that the SAMPL technique scores a large number of loci in the same reaction and thus has a high multiplex ratio. SSR primers, on the other hand, are locus-specific and generally each amplifies a single locus (only rarely, an SSR primer pair may amplify two or three loci), thus having a low multiplex ratio. Therefore a SAMPL reaction is actually equivalent to a large number of SSR reactions. Even a multiplex fluorescently labelled automated SSR reaction will use a maximum of 5–6 primer pairs, and will not be able to detect the level of polymorphism obtained with a solitary SAMPL reaction. Thus, among the microsatellite-based technologies, SAMPL is certainly superior to SSR for studies where locus specificity is not required. In a crop like bread wheat, with a large genome size, it is also superior to AFLP since it gives fewer bands to be conveniently scored and also gives a relatively higher frequency of codominant markers.

Genetic diversity

The parameters like polymorphic information content (PIC) and marker index (MI) are generally used to evaluate a marker system for its ability to detect high levels of DNA polymorphism in an analysis of genetic diversity. The average PIC value of SAMPL primers in the present study in bread wheat was 0.221, which is based on both the frequency of each polymorphic band and the proportion of polymorphic bands in an assay. In another study in bread wheat, mean reported PIC values, when multiplied with a proportion of polymorphic bands (to make them comparable with the present values), turn out to be 0.21 for SSR, 0.08 for RFLP and 0.07 for AFLP (Bohn et al. 1999). In three other earlier studies on bread wheat, PIC values for SSR ranged from 0.23 to 0.79 (Roder et al. 1995), 0.29 to 0.79 (Plaschke et al. 1995) and 0.21 to 0.90 (Prasad et al. 2000). These values will not change significantly, even if the proportion of polymorphic bands is taken into consideration (usually >70–90% SSR loci are polymorphic when studied over a large number of genotypes). Thus the present PIC value for SAMPL is much higher than those reported earlier for RFLP and AFLP, but are either comparable, or lower, than those for SSR. However, such a comparison of a single SAMPL reaction with a number of SSR reactions may not be justified. In view of this, to make the results of SAMPL and SSR comparable, diversity indices were also calculated by using phenotypes rather than individual bands, as done in an earlier study involving AFLP (Russel et al. 1997). This gave a very high value for PIC (0.979), suggesting that individual reactions of SAMPL are extremely informative when compared to individual reactions of SSRs. However, for individual SAMPL markers (individual bands), the average PIC value of 0.264 obtained during the present study is slightly higher than the mean PIC value of 0.20, that was reported for AFLP markers in an earlier study in rye grass (Roldan-Ruiz et al. 2000). This suggests that, although at the individual band level the degree of polymorphism in SAMPL does not differ very much from that reported for AFLP, a high proportion of polymorphic bands led to a value of PIC for the primers, which is much higher than those reported earlier for RFLP or AFLP.

The MI value of SAMPL markers in the present study was 9.61, which is also higher than the value of 3.41 for AFLP obtained in an earlier study in bread wheat (Bohn et al. 1999). These values, when compared with those available for single-locus multiallelic marker systems like SSR prove to be much higher. A low value of MI available in this earlier study involving AFLP was due to a low proportion of polymorphic bands available in this earlier study (0.21 against 0.49 in the present study). An MI value of 6.14 that is intermediate between the above two contrasts was also available for AFLPs in soybean (Powell et al. 1996). In view of the above, when both PIC and MI values are considered together, SAMPL proves to be the most informative marker system for bread wheat.

While evaluating SAMPL for detecting polymorphism, we need to recognize, that all SAMPL bands in a reaction are not polymorphic, as may sometimes be true in SSRs (Prasad et al. 2000). This may be attributed to one or more of the following reasons: (1) lack of polymorphism at some of the compound SSR loci (particularly the compound microsatellites) amplified through SAMPL, (2) a low abundance of polymorphisms in restriction fragments associated with compound microsatellites used in SAMPL, or (3) poor resolution of the polymorphism due to differences of only a single or few repeats. However, a large number of bands or loci available to detect polymorphism in each single SAMPL reaction more than offsets the apparent low polymorphism available at individual band level in this technology, thus making SAMPL technology extremely useful.

The results of the present study were also evaluated for their utility in genotype identification. In this connection, it may be recalled that during the present study SAMPL-6 gave unique patterns in 47 genotypes and

SAMPL-7 gave unique patterns in 49 genotypes. Of these genotypes, as many as 42 gave unique patterns with each of the two SAMPL reactions and both reactions together distinguished 54 of the 55 genotypes. These results suggest that, due to a high multiplex ratio, it is possible to distinguish a fairly large number of wheat cultivars from each other using SAMPL technology.

The mean GS value of 0.65, obtained from 1,485 different pairs derived from 55 genotypes, was much higher than the mean GS value of 0.23 (range=0.05–0.88), obtained using 20 SSRs in an earlier study with the same material (Prasad et al. 2000). Slightly lower values that were reported earlier using an entirely different set of genotypes included 0.57 (21 SSR), 0.65 (117 RFLP probes in combination with five enzymes, giving 470 bands) and 0.61 (16 AFLP primer pairs giving 559 bands) (Bohn et al. 1999). In the present study, a relatively higher value of GS from SAMPL data may be attributed at least partly to the lack of resolution of locus-specific allelic diversity at individual loci in this technology. Low GS values with SSR were also obtained in an earlier study where, using an entirely different material, 23 SSRs gave an average GS value of 0.31 (Plaschke et al. 1995). An unusually high mean GS value (0.81) was obtained in a solitary earlier study of bread wheat, where 45 genotypes were screened using 38 STS markers (Chen et al. 1994), and an unusually low mean GS value (0.18) was obtained using 124 diverse genotypes and a much bigger set of 119 RFLP probes (Paull et al. 1998). In these different studies on genetic diversity in bread wheat undertaken using a variety of molecular markers, the variation in GS coefficient values may be attributed either to the differences in the number/nature of genotypes used or to the differences in the probes/primers employed.

The results of genetic diversity based on the dendrogram derived using GS values and the PCA analysis conducted directly from binary matrix data were in agreement to a great extent (Figs. 2 and 3). The highest GS value of 0.96 was observed between E4229 (from Spain) and E4813 (from Kenya), as well as between E3839 (from Columbia) and E3901 (from Equador), suggesting that the two genotypes in each of these two pairs are closely related. This corresponds well with their positions in the dendrogram as well as in the PCA plot (Figs. 2 and 3). On the basis of available information on pedigree (adequate data on pedigree was not available to work out coefficients of coancestry, *f*), and a high GS value, it seems that they may be ancestrally related in the remote past. In contrast to the above, the pair with genotypes E288 (from Poland) and E3389 (from Lebanon) had the lowest GS value (0.35). The positions of these two genotypes in the dendrogram, as well as in the PCA plot, also suggests that they are diverse. Therefore, they represent a pair of diverse genotypes, representing potential parents for a hybridization programme. In both, cluster analysis and PCA analysis, however, genotypes E3840 from Columbia and E2055 from Poland are unique with respect to other genotypes. The GS value between these two genotypes is 0.457, which is not the lowest, suggesting that these are not so diverse with respect to each other as they are with respect to the other genotypes. Thus the information on genetic similarities (all 1,485 values of GS are not provided), as well as the cluster and PCA analyses collected during the present study, when used together should prove useful for inferring relationships between the genotypes and for the identification of diverse genotypes for plant breeding.

Gene tagging

The five SAMPL primer pairs that were used for gene tagging in the present study gave a fairly high level of polymorphism (37.83% to 62.50%). When considered together for all the five primer pairs, the polymorphic bands were 47.6% for GPC, 51.9% for PHST and 44.7% for GW. In contrast to this, when 340 locus-specific SSR primers were earlier utilized for the same three pairs of parents, 25.1% polymorphic primers were available for GPC (Prasad et al. 1999; Harjit-Singh et al. 2001), 23.9% polymorphic primers were available for PHST (Roy et al. 1999) and 18.2% polymorphic primers were available for GW (Varshney et al. 2000). In case of PHST, a set of 138 STS primers were also tried in an earlier study, where 21.7% primers detected polymorphism (Roy et al. 1999). Thus it is evident that the polymorphism between the same pairs of genotypes detected by SAMPL is 2–3 times the polymorphism detected by locus-specific SSR or STS primers.

Bulked segregant analysis in the present study also allowed the detection of associations between SAMPL markers and each of the three traits, including GPC, PHST and GW (Figs. 4a, b and c; Table 2). Six SAMPL markers were found to be associated with GPC (*XccuES*$_6$*400*, *XccuES*$_6$*390*, *XccuES*$_6$*365*, *XccuES*$_6$*350*, *XccuES*$_6$*325*, *XccuES*$_6$*300*), four SAMPL markers were found to be associated with GW (*XccuES*$_6$*355*, *XccuES*$_6$*340*, *XccuES*$_6$*330*, *XccuES*$_6$*320)* and seven SAMPL markers were found to be associated with PHST (*XccuES*$_6$*430*, *XccuES*$_6$*410*, *XccuES*$_6$*370*, *XccuES*$_6$*345*, *XccuES*$_6$*280*, *XccuES*$_6$*150*, *XccuES*$_6$*110*). These associated SAMPL markers can be converted into locus-specific primers, which would require the elution of associated bands, followed by cloning and sequencing. This is being done in our laboratory in an extended study aimed at developing efficient markers for marker-aided selection.

Conclusion

The present study in bread wheat has demonstrated that the microsatellite-based marker system SAMPL can be efficiently and effectively utilized for a variety of studies involving the detection of DNA polymorphism. It can certainly be used for a study of genetic diversity and for the development of markers for marker-aided selection.

However, studies need to be conducted to find out if the diversity patterns obtained by this marker system differ from the diversity patterns obtained with other marker systems and/or also from the diversity patterns obtained using morphological traits. In order to evaluate their utility, it may also be necessary to determine the relative frequencies of co-dominant markers in AFLP and SAMPL technologies. Further, the markers identified through SAMPL to be associated with economic traits need to be converted into user-friendly PCR-based locus-specific primers for plant breeding.

Acknowledgements Financial support from the Department of Biotechnology, Government of India, the National Agriculture Technology Programme, ICAR, New Delhi and the Council of Scientific and Industrial Research, Government of India is gratefully acknowledged. Thanks are also due to Dr. K.V. Bhatt, National Bureau of Plant Genetic Resources, New Delhi and Dr. M. Laxmikumaran, Tata Energy Research Institute, New Delhi for their help in analyses of data.

References

Bohn M, Friedrich Utz H, Melchinger AE (1999) Genetic similarities among winter wheat cultivars determined on the basis of RFLPs, AFLPs and SSRs and their use for predicting progeny variance. Crop Sci 39:228–237

Chen HB, Martin JM, Lavin M, Talbert LE (1994) Genetic diversity in hard red spring wheat based on sequence-tagged-site PCR markers. Crop Sci 34:1629–1632

Gupta PK, Varshney RK (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica 113:163–185

Gupta PK, Varshney RK, Sharma PC, Ramesh B (1999) Molecular markers and their applications in wheat breeding. Plant Breed 118:369–390

Harjit-Singh, Prasad M, Varshney RK, Roy JK, Balyan HS, Dhaliwal HS, Gupta PK (2001) STMS markers for grain protein content and their validation using near-isogenic lines in bread wheat. Plant Breed 120:1–6

Jaccard P (1908) Nouvelles recherches sur la distribution florale. Bull Soc Vaud Sci Nat 4:223–270

Michelmore RW, Paran I, Kesseli RV (1991) Identification of markers linked to disease resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions using segregating populations. Proc Natl Acad Sci USA 88:9828–9832

Morgante M, Olivieri AM (1993) PCR-amplified microsatellites as markers in plant genetics. Plant J 3:175–182

Morgante M, Vogel J (1994) Compound microsatellite primers for the detection of genetic polymorphisms. U S patent application no. 08/326456

Paglia G, Morgante M (1998) PCR-based multiplex DNA fingerprinting techniques for the analysis of conifer genomes. Mol Breed 4:173–177

Paull JG, Chalmers KJ, Karakousis A, Kretschmer JM, Manning S, Langridge P (1998) Genetic diversity in Australian wheat varieties and breeding material based on RFLP data. Theor Appl Genet 96:435–446

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

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol Breed 2:225–238

Prasad M, Varshney RK, Kumar A, Balyan HS, Sharma PC, Edwards KJ, Singh H, Dhaliwal HS, Roy JK, Gupta PK (1999) A microsatellite marker associated with a QTL for grain protein content on chromosome arm 2DL of bread wheat. Theor Appl Genet 99:341–345

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

Roder MS, Plaschke J, Konig SU, Borner A, Sorrells ME, Tanksley SD, Ganal MW (1995) Abundance, variability and chromosomal location of microsatellites in wheat. Mol Gen Genet 246:327–333

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

Roldan-Ruiz I, Dendauw J, Van Bockstaele E, Depicker A, De Loose M (2000) AFLP markers reveal high polymorphic rates in ryegrasses (*Lolium* spp.). Mol Breed 6:125–134

Roy JK, Prasad M, Varshney RK, Balyan HS, Blake TK, Dhaliwal HS, Singh H, Edwards KJ, Gupta PK (1999) Identification of a microsatellite on chromosome 6B and a STS on 7D of bread wheat showing association with preharvest sprouting tolerance. Theor Appl Genet 99:336–340

Russell J, Fuller J, Young G, Thomas B, Taramino G, Macaulay M, Waugh R, Powell W (1997) Discriminating between barley genotypes using microsatellite markers. Genome 40:442–450

Saghai Maroof MA, Soliman KM, Jorgensen RA, Allard RW (1984) Ribosomal DNA spacer length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proc Natl Acad Sci USA 81:8014–8019

Sneath PHA, Sokal RR (1973) Numerical taxonomy. Freeman, San Francisco

Varshney RK, Prasad M, Roy JK, Kumar N, Harjit-Singh, Dhaliwal HS, Balyan HS and Gupta PK (2000) Identification of eight chromosomes and a microsatellite marker on 1AS associated with QTLs for grain weight in bread wheat. Theor Appl Genet 100:1290–1294

Vivek BS, Simon PW (1999) Linkage relationships among molecular markers and storage root traits of carrot (*Daucus carota* L. ssp. *sativus*). Theor Appl Genet 99:58–64

Vos P, Hogers R, Bleeker R, Reijans M, van de Lee T, Homes M, Frijters A, Pot J, Peleman J, Kupier M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414

Weissenbach J, Gyapay G, Dib C, Vignal A, Morissette J, Millasseau P, Vaysseix G, Lathrop M (1992) A second-generation linkage map of the human genome. Nature 359:794–801

Witsenboer H, Vogel J, Michelmore RW (1997) Identification, genetic localization, and allelic diversity of selectively amplified microsatellite polymorphic loci in lettuce and wild relatives (*Lactuca* spp.). Genome 40:923–936