

I. Bertin · J. H. Zhu · M. D. Gale

SSCP-SNP in pearl millet—a new marker system for comparative genetics

Received: 9 December 2004 / Accepted: 24 February 2005 / Published online: 5 April 2005
© Springer-Verlag 2005

Abstract A considerable array of genomic resources are in place in pearl millet, and marker-aided selection is already in use in the public breeding programme at ICRISAT. This paper describes experiments to extend these publicly available resources to a single nucleotide polymorphism (SNP)-based marker system. A new marker system, single-strand conformational polymorphism (SSCP)-SNP, was developed using annotated rice genomic sequences to initially predict the intron-exon borders in millet expressed sequence tags (ESTs) and then to design primers that would amplify across the introns. An adequate supply of millet ESTs was available for us to identify 299 homologues of single-copy rice genes in which the intron positions could be precisely predicted. PCR primers were then designed to amplify approximately 500-bp genomic fragments containing introns. Analysis of these fragments on SSCP gels revealed considerable polymorphism. A detailed DNA sequence analysis of variation at four of the SSCP-SNP loci over a panel of eight inbred genotypes showed complex patterns of variation, with about one SNP or indel (insertion-deletion) every 59 bp in the introns, but considerably fewer in the exons. About two-thirds of the variation was derived from SNPs and one-third from indels. Most haplotypes were detected by SSCP. As a marker system, SSCP-SNP has lower development costs than simple sequence repeats (SSRs), because much of the work is *in silico*, and similar deployment costs and through-put potential. The rates

of polymorphism were lower but useable, with a mean PIC of 0.49 relative to 0.72 for SSRs in our eight inbred genotype panel screen. The major advantage of the system is in comparative applications. Syntenic information can be used to target SSCP-SNP markers to specific chromosomal regions or, conversely, SSCP-SNP markers can be used to unravel detailed syntenic relationships in specific parts of the genome. Finally, a preliminary analysis showed that the millet SSCP-SNP primers amplified in other cereals with a success rate of about 50%. There is also considerable potential to promote SSCP-SNP to a COS (conserved orthologous set) marker system for application across species by more specifically designing primers to precisely match the model genome sequence.

Introduction

Pearl millet [*Pennisetum glaucum* (L.) R.Br.] is one of the most drought-tolerant cereals and is widely grown in the marginal semi-arid regions of sub-Saharan Africa and in India. This important grain crop is grown as a protein-rich source of food for humans and as a forage crop for livestock. Although pearl millet is regarded as an “orphan” crop, the development of molecular marker systems and molecular marker-based genetic maps was initiated early, in 1990, within a Department for International Development (DfID)-funded programme involving several UK laboratories and breeders at the International Crop Research Institute for the Semi-Arid Tropics (ICRISAT), the Consultative Group on International Agricultural Research (CGIAR) centre near Hyderabad in southern India. As a result, millet breeders at ICRISAT have been among the first to use marker-aided selection (Hash et al. 2003) for developing new commercially available open-pollinated and hybrid cultivars.

Early work concentrated on the development of the first marker-based genetic maps using restriction

Electronic Supplementary Material Supplementary material is available for this article at <http://dx.doi.org/10.1007/s00122-005-1981-0>

Communicated by D.A. Hoisington

I. Bertin · J. H. Zhu · M. D. Gale (✉)
John Innes Centre, Norwich Research Park,
Colney, Norwich,
NR4 7UH, UK
E-mail: mike.gale@bbsrc.ac.uk
Tel.: +44-1603-450599
Fax: +44-1603-450045

fragment length polymorphism (RFLP) markers (Liu et al. 1994). This was quickly followed by the development and mapping of simple sequence repeats (SSRs), of which some 200 are now available (Qi et al. 2001, 2004). From this base the other genetic resources available to millet geneticists and breeders expanded to include comparative maps, of which the first were with rice (*Oryza sativa*) and foxtail millet (*Setaria italica*) (Devos et al. 2000) but pearl millet was soon included the entire grass syntenic synthesis (Devos and Gale 2000), a bacterial artificial chromosome (BAC) library (Allouis et al. 2001) and the first expressed sequence tags (ESTs) (R.N. Mishra et al., unpublished). Today pearl millet is one of the best supported among the minor cereals. This support does not, however, as yet extend to single nucleotide polymorphism (SNP) markers, the object of the present study.

The most common class of polymorphism within different plant species is the SNP. In maize (*Zea mays*), for example, the frequency of nucleotide change among varieties is high, at around one polymorphism per 31 bp in non-coding regions and one polymorphism per 124 bp in coding regions. Insertions and deletions (indels) are also frequent in non-coding regions (one per 85 bp), but rare in coding regions (Ching et al. 2002). SNP frequencies in more conserved crop species may be much lower. For example, a baseline figure for comparing hexaploid bread wheat (*Triticum aestivum*) varieties was established at only one in 1,000 bp, again with more variation in introns than in exons (Bryan et al. 1999).

Various SNP discovery and detection methods are available (Gupta et al. 2001), and the objective of this investigation was to develop a co-dominant and robust system that was at least moderately high throughput, amenable to multiplexing and transferable to breeding laboratories. The strategy chosen was to exploit the expectation of higher SNP and indel levels in introns relative to exons. Briefly, millet ESTs were aligned with annotated rice gene sequences to predict the location of introns, then PCR primers were designed to amplify across introns. The PCR products were then separated by single-strand conformational polymorphism (SSCP) electrophoresis. This provided a relatively polymorphic marker system, termed SSCP-SNP, that will find use in any organism and which has quite different properties and applications from SSRs.

Materials and methods

Pearl millet genotypes

Eight pearl millet inbred lines were used as a genotype panel to assay polymorphism levels: ICMP 451, 81B, 841B, 863B, PT 732B, P1449-2, ICMP 85410 and LGD 1-B-10, all supplied by ICRISAT. These lines are the parents of several of our laboratory mapping populations and were chosen to provide either maximum

levels of polymorphism or variation at key target genes for adaptation.

EST analysis and primer design

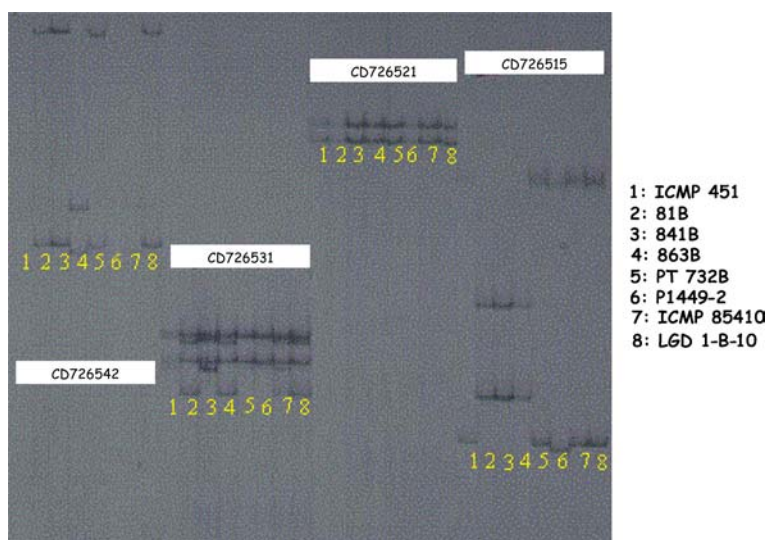
Pearl millet ESTs were obtained from the National Centre for Biotechnology Information (NCBI) and used to query the NCBI database using BLASTN. ESTs showing good homology with rice single-copy genes were selected for further analysis because it was intended, following marker development and mapping, to use the new markers to more closely define the synteny relationship between the pearl millet and rice genomes. Alignment of the selected ESTs with rice genomic sequences from the rice genome automated annotation system (RiceGAAS) available at <http://ricegaas.dna.affrc.go.jp/rgadb/> provided information for predicting intron-exon boundaries in the millet EST. Primer pairs were designed to amplify across intron regions and to give amplified fragments in the 200- to 500-bp range.

PCR and SSCP protocol

Primer pairs to the 299 selected ESTs were tested for PCR amplification and polymorphism between the eight inbred genotypes on SSCP gels. Amplification was conducted in a 20- μ l volume containing 100 ng genomic DNA, 1 μ M of each primer, 1.25 mM dNTPs, 1 \times PCR buffer (Roche, Indianapolis, Ind.) and 0.4 U *Taq* DNA polymerase (Roche). After an initial denaturation at 94°C for 1 min, 30 amplification cycles were performed—94°C for 1 min (denaturation), 56°C for 1 min (annealing at an optimal temperature) and 72°C for 1 min (extension). The PCR products were denatured and separated on 310 \times 380 \times 0.4-mm SSCP gels using a mutation detection enhancement (MDE) gel solution (Martins-Lopes et al. 2001). The gel mix was made in a 60-ml total volume containing a final concentration of 0.5 \times gel solution (Cambrex Bio-science Rockland, Rockland, Me.) and 0.6 \times TBE buffer and polymerized by the addition of 0.24 ml of 10% ammonium persulphate and 24 μ l of tetramethylenediamine (TEMED). Fragments were electrophoresed for 16 h at a constant power of 8 W at room temperature and then silver-stained as described by Bassam et al. (1991).

Initial screening of the genotype panel was carried out with single-loaded gels (see Fig. 1). Individual markers were then classified as having band clusters running at the “top”, “middle” or “bottom” of the gel. In subsequent assays—for example, in mapping gels—this information was used to co-load two amplification products in single tracks. Following this protocol, we have added 96 new SSCP-SNP loci to the millet consensus map and re-assessed the syntenic alignment of the pearl millet and rice genome. These results, which are extensive, will be published separately.

Fig. 1 A portion of an SSCP-SNP panel screen gel. Three markers show polymorphism: PSMS12CD726542 shows two SSCP alleles and three amplification failure tracks; PSMS13CD726531 shows two alleles at one locus and amplification at a second monomorphic locus; CD726521 is monomorphic; and PSMS15CD726515 shows three alleles at a single locus



Nomenclature

The nomenclature adopted for SSCP-SNP markers for pearl millet is of the form PSMS90CD725138 (Pyruvate decarboxylase) for the amplicon and *Xpsms90CD725138* (Pyruvate decarboxylase) for the locus. *X* indicates a DNA marker (following the wheat nomenclature), *ps* identifies a John Innes Centre marker; *m* indicates pearl millet; *s* indicates a SNP marker; 90 is a unique PSMS marker number, which relates to specific forward and reverse PCR primers; CD725138 is the Genbank accession number of the source EST; (Pyruvate decarboxylase) indicates putative function. The function identifier is optional. In some circumstances, the locus name may be abbreviated to simply *Xpsms90*.

Results and discussion

The SSCP-SNP marker system

The identification of likely single-copy markers was achieved entirely in silico. A total of 1,900 pearl millet EST sequences were screened and duplicates removed. Of the 650 (34%) with good homology to rice gene sequences, 299 (16%) were homologous to single-copy genes. Primer pairs were designed over predicted intron locations for these 299 ESTs, and 77% amplified well across the genotype panel. Primer failure was probably due to sequence variation in the primer sequence itself in our panel genotypes or, less likely, to millet introns that were too long to PCR effectively. The precise cause of individual “null” tracks, as in Fig. 1, is not known because the initial panel screens were rarely repeated. Although intron positions were well predicted by the rice gene annotation, intron length varied considerably between rice and millet in amplicons that were fully sequenced (see below). The 231 primer pairs provided us with 102 markers revealing polymorphism at 106 loci.

Most millet SSCP-SNP markers selected in this way addressed a single map location. Fulton et al. (2002), during the development of the tomato (*Lycopersicon esculentum*) conserved orthologue set (COS) markers, similarly scanned tomato ESTs against the arabidopsis sequence. In the majority of cases, those tomato markers which hybridised with a single arabidopsis BAC detected single or low-copy genes in most of the plant species tested; for example, sunflower, lettuce, melon, alfalfa, rice and cotton.

DNA separation on SSCP gels is based on both size and conformation. As the conformation cannot be known in advance, the precise relative gel positions cannot be predicted. Nevertheless, the mobility of the typical two-band SSCP profile for single-copy amplifications is consistent over runs and, once assessed, can be used as the basis for post-PCR multiplexing by co-loading.

The polymorphism information content (PIC), a measure of allelic diversity at a locus, was estimated for each of the polymorphic SSCP-SNP loci detected among the eight genotype panel in this study. The mean PIC value was 0.49. A comparable value for 39 polymorphic SSRs (Qi et al. 2004) on the same pearl millet panel was 0.72.

Table S1 in the electronic supplementary material (ESM) shows the Genbank numbers, forward and reverse primers for each marker, together with SSCP allelic genotypes for each of the eight cultivars in the panel. Additional data include the PIC value over the eight genotypes, the SSCP gel position and the putative gene function from the gene function selector in Rice-GAAS. Primer information is also provided for the monomorphic markers in Table S2 of the ESM.

The relationship between haplotype and SSCP phenotype

Four SSCP-SNP loci were examined in detail by sequencing the amplification products of each of the

eight inbred genotypes. The aim here was to quantify the SNP and indel variation present, to observe the relative frequencies of variation in exon and intron sequences and to attempt to gauge how much of this variation was reflected in the SSCP band patterns. Two of these analyses are shown in Fig. 2, while data on the two other loci are shown in Fig. S1 of the ESM.

All SSCP-SNP markers in this study had their PCR primers designed to be in conserved regions and to amplify across the assumed more polymorphic intron regions. The closest rice homologue was used to infer the intron position and thereby help primer design. Over the four loci one had an intron of exactly the same length in rice and millet, one was shorter in millet and two were longer. The largest difference was in PSMS97CD726001 with a 575-bp millet intron, while the rice homologue contains only a 338-bp intron. In no case was there any sequence homology between rice and millet in the intron regions.

Over the 1,530 bp observed in detail, the frequency of SNP to indel was 1 per 59 bp in the introns and 1 per

714 bp in the exons, which clearly justifies the intron-targeted marker strategy. About two-thirds of the variation was observed as SNPs and one-third as indels. These results are consistent with those reported by Ching et al. (2002) in maize.

The SSCP phenotype and variation in DNA sequence over the eight panel genotypes at four loci are shown in Figs. 2 and ESM S1. The four loci were all polymorphic and cannot be said to have been chosen at random. Nevertheless, the level of variation in DNA sequence was high. PSMS22CD726311 (Fig. 2) showed three haplotypes derived from two long indels with two associated SNPs and one independent SNP. All of the variation was in the intron. PSMS29CD726044 (Fig. 2) is very complex and includes an extremely polymorphic AC/TC/TGA chimeric microsatellite together with another microsatellite, three SNPs and an indel, again all in the intron region. PSMS15CD726515 (ESM Fig. S1) has three haplotypes deriving from one SNP in the exon region and one indel in the intron. PSMS97CD726001 (ESM Fig. S1) has two haplotypes involving an indel and two associated SNPs.

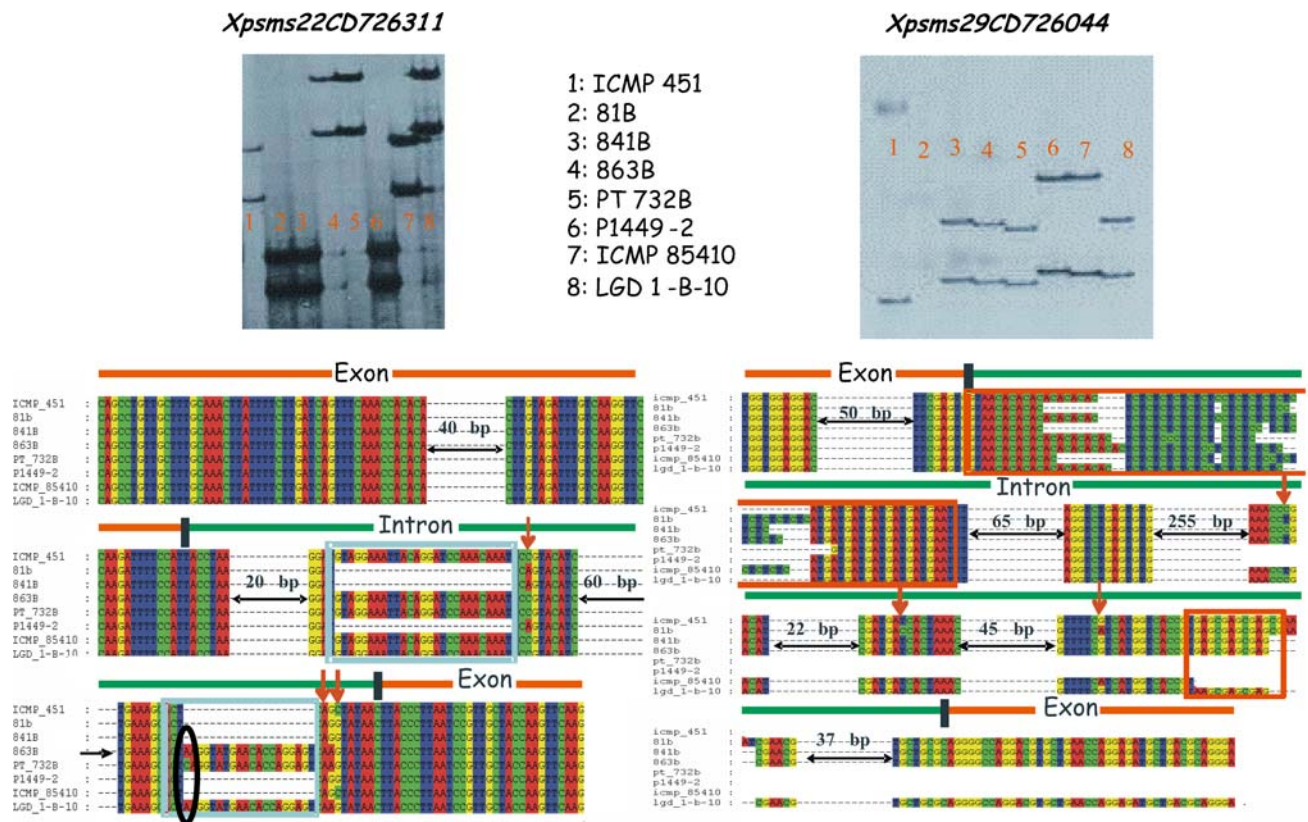


Fig. 2 SSCP gel profiles and sequence data at two loci, *Xpsms22CD726311* and *Xpsms30CD726044*, over the genotype panel (1–8). SNPs are identified by red arrows, indel regions are boxed in blue and microsatellite regions are boxed in red. SNPs undetected on SSCP gels are circled in black. To enable a concise display some regions within which no variation was detected over the eight panel genotypes are indicated by black arrows with the sequence length indicated above the arrows. *Xpsms22CD726311* displays two large indels and three associated SNPs, which partition the eight panel

genotypes into three haplotypes, all recognised by the SSCP phenotype. A further C/A SNP in PT 732B is, however, not detected by SSCP. *Xpsms30CD726044* is exceptionally polymorphic on the SSCP gel. This turns out to be mainly due to a chimeric AC/TC/ATG microsatellite embedded in the intron sequence. Three more SNPs and another small microsatellite, GAGC, are also present in the intron sequence. No variation is observed in the 105-bp exon region

It is clear that most of the DNA haplotype variation is reflected in the SSCP phenotype. All of the variants at *Xpsms15CD726515* and *Xpsms97CD726001* and the two major haplotypes at *Xpsms22CD726311* gave distinct SSCP band patterns. Only one SNP at *Xpsms22CD726311* was not detected by SSCP. Clearly at some loci—for example, *Xpsms29CD726044*—the indel/SNP haplotypes can be so complex that it is impossible to determine whether all polymorphism has been translated to band pattern variation. Martins-Lopes et al. (2001) also tested the efficiency of SNP detection by SSCP gels. In their experiments, four amplified fragments, differing by one to six SNPs, were investigated using MDE-based SSCP gels and all gave distinct SSCP band patterns. Thus, SSCP gels are probably an efficient way to identify the polymorphism present in the fragments amplified by the 299 primer pairs tested in this study.

SSCP-SNP as a COS marker system

The main aim of the present study was to develop markers specifically for pearl millet with maximum comparative utility. Nevertheless, it is of interest to determine whether the SSCP-SNP markers will find direct application in other cereals. One can be sure that, if the primers amplify at all, the amplicon will contain an intron in any species. A preliminary screen of 15 pairs of primers with rice, wheat, barley (*Hordeum vulgare*) and finger millet (*Eleusine coracana*) genomic DNA showed about 50% success. Finger millet amplified with nine primers, while with barley only three primers amplified well.

An *in silico* comparison of millet and rice over the primer sequences showed that, except where the match was perfect, it was difficult to predict whether the millet primers would amplify rice genomic DNA. However, the analysis did show clearly that the success rate could have been promoted to almost 100%, with perfect matches in nearly all cases for both primers, had the markers been designed with COS markers in mind. It is also probable that markers designed in this way would amplify more consistently over species.

Conclusions

SSCP-SNP is a new PCR gel-based, co-dominant marker system applicable to all organisms. The system has the same multiplex potential as SSRs. Although mobility on SSCP gels is not well predicted by molecular weight alone, the position on the gel of individual SSCP phenotypes may be classified *post hoc* and this information used in subsequent assays for co-running multiple markers in the same track. In this study, individually amplified reactions were loaded at different times in the

same track. The use of fluorescent-labelled primers (Scholl et al. 2001; Grieu et al. 2003) will increase multiplex possibilities further.

SSCP-SNP should find application in breeding programmes as an alternative to SSRs. The same equipment and laboratory techniques are needed, the only difference being that SSCP uses an MDE gel matrix rather than the denaturing polyacrylamide gels used to resolve SSRs. While detection costs are similar to SSR, SSCP-SNP development costs are less than for SSRs because database mining is far less costly than wet-bench library construction. The system is, of course, dependant on an adequate supply of pre-existing ESTs. SSCP-SNP is, however, less polymorphic, giving mean PIC values of 0.49 relative to 0.72 for SSRs tested on the same genotype panel.

Differences in the SSCP phenotype reflect most of the small changes (SNPs and indels) in the DNA sequence of the amplified fragments. However, getting to the haplotype underlying any SSCP pattern is not trivial since individual amplicons must be cloned and sequenced at least twice to guard against sequencing errors.

The great advantage of the system comes in relation to synteny. As ESTs with known homology to single-copy genes in fully sequenced model species, in this case rice, can be selected as the source of SSCP-SNP markers, every mapped locus provides information on synteny between the genomes of the target and that of the model. Once some knowledge of the syntenic relationships has been gained, as is already available for pearl millet and rice (Devos et al. 2000), this information can be used to target SSCP-SNP markers to specific regions simply by choosing ESTs with homology to genes in the syntenic region of the model. Example applications are to identify new markers for specific genes that breeders may want tagged for marker-aided selection or to increase marker density in genome regions where more detailed knowledge of synteny is required. The two loci described in Fig. 2 are cases in point. The rice homologue to CD726311 lies near the centromere on rice chromosome 2S (R2S), and known synteny would predict that *Xpsms22* should provide an additional marker in the telomeric region at the bottom of millet linkage group 2 (LG2), which is exactly where it maps. *Xpsms29* maps to the bottom region of millet LG1 and provides the first confirmation of synteny between that region and a small segment of R11S. Using the same arguments, SSCP-SNP should also find considerable application for fine mapping as a prelude to gene isolation.

Acknowledgements This manuscript is an output from projects funded by the UK DfID for the benefit of developing countries. The views expressed are not necessarily those of DfID. We also wish to acknowledge Prof. M.K. Reddy's group at the Plant Molecular Biology Laboratory, International Centre for Genetic Engineering and Biotechnology, New Delhi for developing the ESTs used in this study.

References

- Allouis S, Qi X, Lindup S, Gale MD, Devos KM (2001) Construction of a BAC library of pearl millet, *Pennisetum glaucum*. *Theor Appl Genet* 102:1200–1205
- Bassam BJ, Caetano-Anollés G, Gresshoff PM (1991) Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal Biochem* 196:80–83
- Bryan GJ, Stephenson P, Collins A, Kirby J, Smith JB, Gale MD (1999). Low levels of DNA sequence variation among adapted genotypes of hexaploid wheat. *Theor Appl Genet* 99:192–198
- Ching A, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, Morgante M, Rafalski AJ (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. *BMC Genet* 3:19
- Devos KM, Gale MD (2000) Genome relationships: the grass model in current research. *Plant Cell* 12:637–646
- Devos KM, Pittaway TS, Reynolds A, Gale MD (2000) Comparative mapping reveals a complex relationship between the pearl millet genome and those of foxtail millet and rice. *Theor Appl Genet* 100:190–198
- Fulton TM, Van der Hoeven R, Eannetta NT, Tanksley SD (2002) Identification, analysis, and utilization of conserved ortholog set markers for comparative genomics in higher plants. *Plant Cell* 14:1457–1467
- Grieu F, Joseph D, Norman P, Iacopetta B (2003) Development of a rapid genotyping method for single nucleotide polymorphisms and its application in cancer studies. *Oncol Rep* 11:501–504
- Gupta PK, Roy JK, Prasad M (2001) Single nucleotide polymorphisms: a new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. *Curr Sci* 80:524–535
- Hash CT, Raj AGB, Lindup S, Sharma A, Beniwal CR, Folkertsma RT, Mahalakshmi V, Zerbini E, Blummel M (2003) Opportunities for marker-assisted selection (MAS) to improve the feed quality of crop residues in pearl millet and sorghum. *Field Crops Res* 84:79–88
- Liu CJ, Witcombe JR, Pittaway TS, Nash M, Busso CS, Hash CT, Gale MD (1994) An RFLP-based genetic map of pearl millet (*Pennisetum glaucum*). *Theor Appl Genet* 89:481–487
- Martins-Lopes P, Zhang H, Koebner R (2001) Detection of single nucleotide mutations in wheat using single strand conformation polymorphism gels. *Plant Mol Biol Rep* 19:159–162
- Qi X, Lindup S, Pittaway TS, Allouis S, Gale MD, Devos KM (2001) Development of simple sequence repeat markers from bacterial artificial chromosomes without subcloning. *Biotechniques* 31:355–362
- Qi X, Pittaway TS, Lindup S, Liu H, Waterman E, Padi FK, Hash CT, Zhu J, Gale MD, Devos KM (2004) An integrated genetic map and a new set of simple sequence repeat markers for pearl millet, *Pennisetum glaucum*. *Theor Appl Genet* 109:1485–1493
- Scholl R, Walker A, Wood BL (2001) Multiplex, fluorescent single-strand conformation polymorphism using stepped polymerase chain reaction primers. *J Biomol Technol* 12:1–3