

Research article

Open Access

## Development and mapping of Simple Sequence Repeat markers for pearl millet from data mining of Expressed Sequence Tags

S Senthilvel, B Jayashree, V Mahalakshmi, P Sathish Kumar, S Nakka, T Nepolean and CT Hash\*

Address: International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, Andhra Pradesh, 502 324, India

Email: S Senthilvel - s.senthilvel@cgiar.org; B Jayashree - b.jayashree@cgiar.org; V Mahalakshmi - maha\_balu@hotmail.com; P Sathish Kumar - ponskumar@gmail.com; S Nakka - sri\_nakka@yahoo.com; T Nepolean - t.nepolean@cgiar.org; CT Hash\* - c.hash@cgiar.org

\* Corresponding author

Published: 27 November 2008

Received: 29 April 2008

BMC Plant Biology 2008, 8:119 doi:10.1186/1471-2229-8-119

Accepted: 27 November 2008

This article is available from: <http://www.biomedcentral.com/1471-2229/8/119>

© 2008 Senthilvel et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** Pearl millet [*Pennisetum glaucum* (L.) R. Br.] is a staple food and fodder crop of marginal agricultural lands of sub-Saharan Africa and the Indian subcontinent. It is also a summer forage crop in the southern USA, Australia and Latin America, and is the preferred mulch in Brazilian no-till soybean production systems. Use of molecular marker technology for pearl millet genetic improvement has been limited. Progress is hampered by insufficient numbers of PCR-compatible co-dominant markers that can be used readily in applied breeding programmes. Therefore, we sought to develop additional SSR markers for the pearl millet research community.

**Results:** A set of new pearl millet SSR markers were developed using available sequence information from 3520 expressed sequence tags (ESTs). After clustering, unigene sequences (2175 singlets and 317 contigs) were searched for the presence of SSRs. We detected 164 sequences containing SSRs (at least 14 bases in length), with a density of one per 1.75 kb of EST sequence. Dinucleotide repeats were the most abundant followed by tri-nucleotide repeats. Ninety primer pairs were designed and tested for their ability to detect polymorphism across a panel of 11 pairs of pearl millet mapping population parental lines. Clear amplification products were obtained for 58 primer pairs. Of these, 15 were monomorphic across the panel. A subset of 21 polymorphic EST-SSRs and 6 recently developed genomic SSR markers were mapped using existing mapping populations. Linkage map positions of these EST-SSR were compared by homology search with mapped rice genomic sequences on the basis of pearl millet-rice synteny. Most new EST-SSR markers mapped to distal regions of linkage groups, often to previous gaps in these linkage maps. These new EST-SSRs are now used by ICRISAT in pearl millet diversity assessment and marker-aided breeding programs.

**Conclusion:** This study has demonstrated the potential of EST-derived SSR primer pairs in pearl millet. As reported for other crops, EST-derived SSRs provide a cost-saving marker development option in pearl millet. Resources developed in this study have added a sizeable number of useful SSRs to the existing repertoire of circa 100 genomic SSRs that were previously available to pearl millet researchers.

## Background

Pearl millet [*Pennisetum glaucum* (L.) R. Br.] is the staple food and fodder crop of millions of poor people living on the most marginal agricultural lands of sub-Saharan Africa and the Indian subcontinent. Indeed, in some of the hottest and driest regions where agriculture is possible in India and Africa, pearl millet is the only cereal that can be grown under dryland conditions and so plays a critical role in food security. However, people living in these regions have not yet benefited much from the current 'biotechnology revolution', or even the 'green revolution' that dramatically increased food grain production on irrigated lands over a generation ago. In recent times, the use of molecular marker technology for the genetic improvement of pearl millet has made some headway, and pearl millet has been elevated to the status of a molecular crop through a series of collaborative projects involving the John Innes Centre (JIC), ICRISAT and their partners supported by the Plant Sciences Research Programme of the UK's Department for International Development (DFID) [1], and the Generation Challenge Programme (GCP) of the Consultative Group on International Agricultural Research (CGIAR).

The first major milestone was achieved in 1993 with the creation of a molecular marker-based genetic linkage map of the pearl millet genome with 181 restriction fragment length polymorphism (RFLP) markers – the marker system of choice in the early 1990s [2]. Now more than 600 molecular markers have been created and mapped for pearl millet, a consensus linkage map has been produced [3], and quantitative trait loci (QTL) for disease resistance [4-8], drought tolerance [9-11], flowering time and grain and stover yield [12], and ruminant nutritional quality of straw [13] have been mapped. These genetic tools for marker-assisted breeding of pearl millet are now in place and available for anyone to use in improving pearl millet hybrids and to extend the economic lifespan of elite hybrid parental lines. However, application of these discoveries is hampered by the limited availability of repeatable, polymorphic PCR-compatible markers for pearl millet. ICRISAT and its partners have successfully demonstrated the use of RFLP markers in the marker-assisted backcross transfer of additional downy mildew resistance into a parental line of popular pearl millet hybrid "HHB 67" and a new version of this hybrid, "HHB 67 Improved" based on the improved parental line, was released for commercial cultivation in 2005 as the first public-bred product of marker-assisted breeding in India [14,15]. However, RFLP markers are too labor-intensive and high cost for applied use, as well as having potential health and environmental hazards. Thus these markers are not considered suitable for large-scale genotyping applications in an applied plant breeding program.

For regular plant breeding applications, PCR-compatible markers based on microsatellites or simple sequence repeats (SSRs) are often considered the most appropriate. SSRs typically provide single-locus markers, which are often co-dominantly inherited and characterized by hypervariability, abundance and reproducibility. Currently, circa 110 SSR markers are available for pearl millet in public domain out of which 65 SSR loci have been mapped [16-20], but a much larger number is required for their application in plant breeding. Therefore, development of additional SSR markers is a valuable objective for the pearl millet research community. In the past, SSRs have been expensive to develop and this has largely limited their application to the more commercially important crops. Enrichment protocols have been used to reduce these costs by focusing sequencing efforts on DNA clones that are likely to contain a particular repeat motif as done by Budak and co-workers. [17].

Recent studies have indicated large numbers of SSRs in coding regions [21,22], which can act as an alternative source of microsatellites. Discovery of microsatellites in Expressed Sequence Tags (ESTs) provides the opportunity to develop SSR markers (EST-SSRs) in a simple and direct way, *i.e.*, by electronic searches (data mining) of EST databases. Exploitation of this source of SSR markers is obviously limited to the species for which EST sequence information is available. In recent times, lots of efforts have gone into generation of substantial expressed sequence tag (EST) databases for plant species. The exploitation of EST databases to develop microsatellite markers was first attempted in rice [23] and has subsequently been reported from many other plant species [24].

EST-SSRs constitute a novel source of markers that are physically associated with coding regions of the genome, and this can enhance the role of genetic markers in germplasm evaluation by enabling assay of the variation in transcribed sequences and genes where function is known. Contrary to EST-SSRs, genomic SSRs are anonymous markers, and the genetic linkage they sometimes exhibit with quantitative trait loci is often only due to a physical linkage and statistical relationship [25]. In addition, EST-SSRs are easily transferable to closely related species [26,27]. As EST-SSR markers target coding regions of the genome, they facilitate the use of the candidate gene mapping approach, thereby facilitating the dissection of complex traits [28]. This marker system offers the possibility of selecting markers according to the biochemical and physiological properties of their gene products in relation to the phenotype [29]. Considering the advantages and utility of EST-SSR markers, we attempted to develop them from the limited number of EST sequences available for pearl millet.

## Results & discussion

### Identification of SSR

The total number of EST sequences available for pearl millet was 3520 at the start of this study. These sequences spanning the length of 1.4 Mb was assembled using CAP3. The CAP3 analysis resulted in identification of 2175 singlets and 317 contigs. The singlets and contig sequences (1.1 Mb) were searched for SSR motifs. There were 164 sequences containing SSRs, out of which 28 had more than one SSR. The total number of SSR motifs identified was 199 including 23 SSRs in compound form. Thus the frequency of SSRs was 6.6 per cent and comparable with other cereals [30] and the density was one SSR in 1.75 kb of EST, which is comparable to sorghum, rice [31] and wheat [21]. However, in the study of Varshney *et al.* [30] SSRs were reported to occur at a frequency of every 4–8 kb in cereals (barley, maize, rice, rye, sorghum and wheat). Varied results could be found in terms of frequency of SSRs in literature because the parameters used for identification of SSRs as well as the number of sequences assayed varied across studies. SSRs have been found in a range of 6–11% of the total EST sequences from several cereal species *viz.*, barley, maize, rye, rice, sorghum and wheat [30]. However, Kantety and co-workers reported comparatively low frequencies of SSR motifs (1.5% in maize to 4.7% in rice) in the EST collections of cereals as they considered SSRs of only 18–20 nucleotides in length [31]. It should be noted that the frequency and density estimates in this study might not reflect the exact picture for pearl millet considering the limited number of EST sequences analysed.

Unlike other crops, di-nucleotide repeats were the most abundant group detected in this pearl millet study. When SSRs of 12 bp in length were considered, tri-nucleotide repeats were more frequent than di-nucleotide repeats in other cereals [31]. In the current study, the relative abundance of di-, tri-, tetra-, penta and hexa-nucleotide repeat motifs was 81 (40%), 53 (27%), 24 (12%), 29 (15%) and 12 (6%), respectively. Among the di-nucleotide motifs, AG/CT was the most abundant (81.5%), which was in line with other cereals [31]. The AG/CT motif can represent codons GAG, AGA, UCU, and CUC in mRNA populations and translate into the amino acids Arg, Glu, Ala, and Leu, respectively. Ala and Leu are present in high frequencies in proteins and this could be one of the reasons for the predominance of AG/CT motifs in EST collections [31]. The least frequent di-nucleotide repeat class was CG (1.23%). In other studies also, CG was found to be the least common motif in most of the cereal species except barley [30,32]. Although AT motifs are reported to be dominant in plant genomes [33-35]; this motif was present only at the rate of 6.17% in the present study. Among the tri-nucleotide repeats, AGC/CGT was the single largest repeat class (26.42%).

Out of 199 SSRs, 141 were Type II microsatellites (having less than 20 bases). In general, it has been observed that repeat numbers and total lengths of SSRs in transcribed regions are relatively small compared to other regions of genome [31,36].

### SSR amplification

Primer pairs could not be synthesized for about 40% of the SSR-containing EST sequences, as the microsatellite was located close to one of the ends of the sequence or flanking sequences were inappropriate for designing high quality primer pairs. Thus, primer pairs were synthesized from 90 unigene sequences only. These primer pairs were used to amplify genomic DNA of 11 pairs of pearl millet mapping population parental lines. Out of 90 primer pairs developed, 15 did not work, 17 produced multiple or non-specific fragments and 58 were functional giving simple PCR products. Most of the primer pairs that failed and the primer pairs that amplified non-scorable products had been designed from EST sequences with compound SSRs or from contig sequences.

Thirty-five percent of the primer pairs designed for pearl millet EST-SSRs did not amplify a product or give simple PCR products. However, a larger percentage of genomic SSR primer pairs have reportedly failed to amplify in pearl millet (55%) [16]. The lack of amplification by some of the EST-SSR markers has been routinely reported in several previous studies in other crops [36-40], which were attributed to primer mismatches, the extension of primers across a splice site or the presence of large introns in the genomic DNA template. The quality of EST sequence is therefore crucial if one is to obtain functional primers with good amplification profiles. Nevertheless the success rate is higher with EST-SSRs than with genomic SSRs. In maize, it has been noted that a higher percentage of primer pairs from enriched genomic libraries failed to give consistent amplification products than from EST-sequence-derived primer pairs in maize [41]. The authors commented that this could be due to SSR candidates being located in repeated or complex sequences in the genomic clones as opposed to lower-copy sequence origins for the EST-derived candidates. In addition, redundancy of clones within the enriched genomic libraries limited the efficiency of identifying novel SSRs.

Non-specific amplifications were a problem for 18% of the primer pairs that we analysed. These primer pairs produced more bands than expected. A similar trend has also been observed in tall fescue and this might be due to amplifications of loci from the duplicated genomic regions [39]. In wheat also, 39% of EST-SSR markers detected multiple loci [38], which is probably due to the high rate of conservation of EST-SSRs across the three

genomes of this species, as well as 25–30% gene duplication [42].

There were some differences observed between the expected and the observed sizes of the amplification products, which has been a regular phenomenon with EST-SSRs [43,26]. Among 58 primer pairs producing good quality amplification product, 9 returned product sizes larger than expected, which could be due to simultaneous amplification of an intron during the PCR. Conversely, 9 primer pairs gave amplification products that were smaller than expected, suggesting the possibility of deletions within the genomic sequences or lack of specificity of some primer pairs, which might have resulted in the amplification of a different copy belonging to the same multi-gene family. In a few cases, the presence of introns and insertions-deletions (in-dels) has been confirmed by sequence analysis [39]. The forward and reverse sequences, annealing temperatures, and expected product sizes for the 58 functional primer pairs are provided (see Additional file 1).

#### **Assessment of polymorphism in mapping population parental pairs**

Out of 58 functional EST-SSR primer pairs, 15 were found monomorphic across the surveyed panel of mapping population parental lines. The average number of alleles detected across the surveyed panel of genotypes for the 43 polymorphic loci was just 3.1. Nineteen primer pairs produced only 2 alleles across this panel. The highest number of alleles (8) was produced by primer pair ICMP3008. As expected, the genomic SSRs tested were comparatively more polymorphic. Out of 18 genomic SSR primer pairs (CTM series), 4 did not amplify and for the remaining 14 functional primer pairs, 10 detected polymorphism in the surveyed panel of parental genotypes. The average number of alleles produced by these 10 genomic SSRs was 5.0. The polymorphism information for EST-SSRs and genomic SSRs across the set of mapping population parental lines are provided (see Additional file 2 and 3).

It has been reported that EST-derived SSRs are less polymorphic than those derived from genomic libraries. However, EST-SSRs represent a unique opportunity to exploit existing sequence information and by-pass creation and sequencing of SSR-enriched (or random) genomic libraries. Perhaps the most important feature of EST-SSRs is that the primer pairs designed for them are more likely to function in distantly related species than are SSR primer pairs derived from genomic libraries [31,44]. This makes EST-SSRs potentially more useful for comparative mapping studies. Once mapped, EST-SSRs provide map locations for the genes that carry them. We therefore propose to use these markers in other less studied crops like finger millet (*Eleusine coracana*) and foxtail millet (*Setaria italica*), and

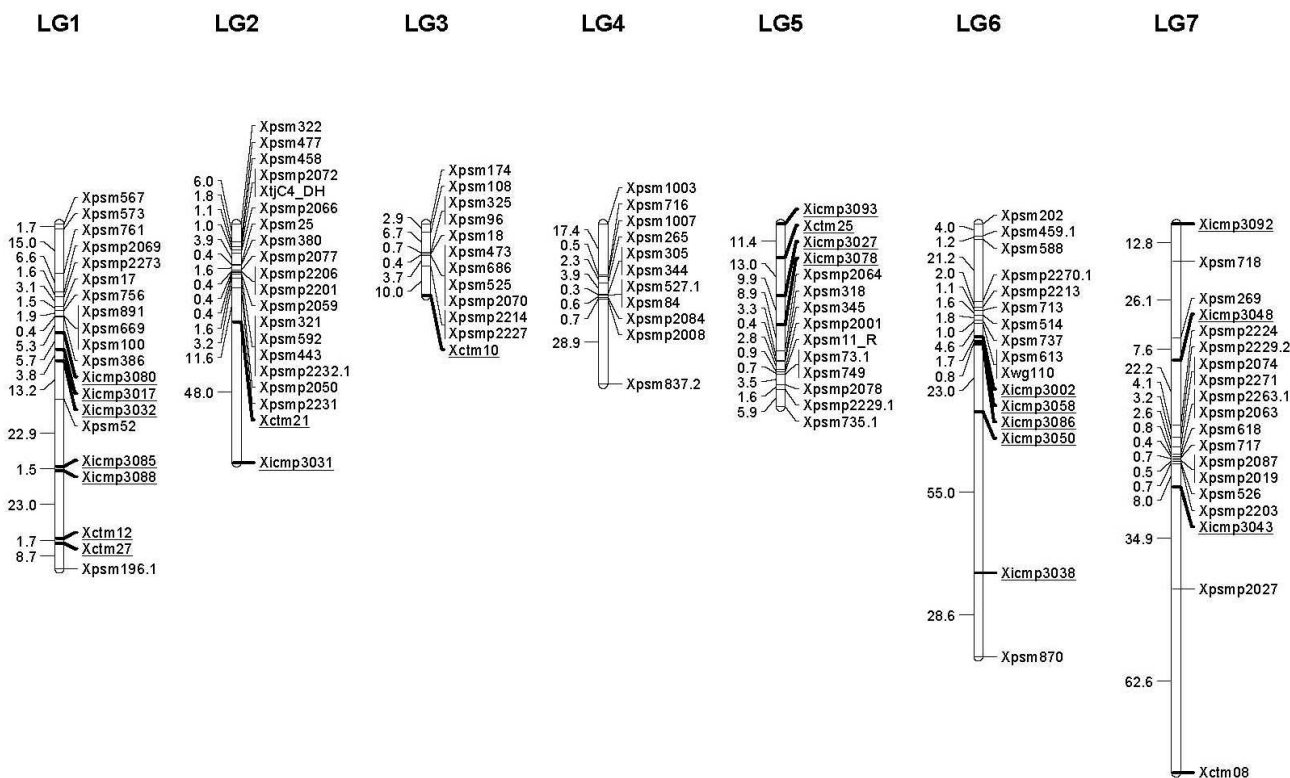
have initiated assessment of the ability of the pearl millet EST-SSR primer pairs described here to amplify product and detect polymorphism in these small millets.

#### **Mapping of SSR markers**

The segregation pattern of loci detected by 17 EST-SSR and 6 genomic SSR primer pairs that were polymorphic between ICMB 841-P3 and 863B-P2 were assessed using 149 F<sub>2</sub> individuals. One of the genomic SSRs, *Xctm08* detected by primer pair CTM08, behaved in dominant fashion and segregated in the ratio of 3:1. The remaining 22 markers segregated in a co-dominant manner. However, 11 of the 17 EST-SSR markers did not segregate in the expected ratio of 1:2:1. Instead, they exhibited skewed distributions towards one parental allele or the other. Such segregation distortion has often been noted in pearl millet. Most of such distortions are cross-specific except that distortion on Linkage Group (LG) 4 was present in three of the four crosses analyzed in an earlier study [16]. In all the crosses of that study, there was an excess of one of the parental alleles in the region spanned by markers *Xpsm265* and *Xpsm364*, suggesting presence of a gene in this region that affects gametophytic or zygotic viability. Previously, segregation distortion was noted on LG 3 and LG 6 of the ICMB 841-P3 × 863B-P2 cross [10,11] with an excess of ICMB 841 alleles. In this study too, 6 out of 11 distorted markers were from LG 6. However the skewness was not only always towards the alleles of female parent ICMB 841, but in some cases favored those of male parent 863B-P2. Out of 6 genomic SSRs, the segregation pattern for one that mapped to LG 3 was distorted.

The 17 EST-SSR markers were mapped to 5 of the 7 pearl millet LGs. The map locations of these 17 new EST-SSRs and 6 genomic SSRs are shown in Figure 1. The characteristic feature of pearl millet linkage maps is the presence of large gaps in distal regions of the chromosomes. Most of the RFLP and SSR markers developed to date have mapped to centromeric regions. Past attempts to develop more markers targeting more distal regions of the pearl millet chromosomes have resulted in identification of few additional markers [16]. The EST-SSRs developed in this study were expected to map to gene-rich portions of the genome, which are expected to be associated with non-centromeric regions of the seven pearl millet chromosome pairs.

The previously constructed framework map for the cross ICMB 841-P3 × 863B-P2 [10,11] had large gaps in LG 1 (between *Xpsm52* and *Xpsm196.1*), LG 2 (between *Xpsm708.1* and *Xpsm322*; and between *Xpsm2237* and *Xpsm708.2*), LG 6 (between *Xwg110* and *Xpsm870*), and LG 7 (between *Xpsm718* and *Xpsm269*; and between *Xpsmp2203* and *Xpsmp2027*). Interestingly, some of the new EST-SSR markers mapped to these gaps. For example,



**Figure 1**  
**Linkage map of ICMB 841-P3 x 863B-P2 showing the map positions of new EST-SSR and genomic SSR markers relative to markers previously mapped in this population. Newly added markers are underlined.**

the 63 cM gap between *Xpsm52* and *Xpsm196.1* on LG 1 was covered by *Xicmp3085*, *Xicmp3088*, *Xctm112* and *Xctm27*.

So far, the shortest linkage group in pearl millet was LG 3. This linkage group was thought to be reasonably complete using comparative mapping as a guide [45]. In this study, one additional marker *Xctm12* was mapped at distal end of LG 3, increasing map length of this LG by 8 cM. The completeness of another short linkage group, LG 5, was not confirmed earlier. In this study, 4 additional markers (*Xicmp3093*, *Xctm25*, *Xicmp3027*, and *Xicmp3078*) were mapped to the top of LG 5 increasing the map length of this LG in this cross to 62 cM from 23 cM. This is in agreement with the earlier prediction that markers would eventually be found that will map 30 cM or more beyond the most distal markers on LG 3 and LG 5 [16,45]. The gap between *Xwg110* and *Xpsm870* in LG 6 was filled with 5 new markers but the distance between these two markers increased drastically, calling into question the positioning of these markers.

The average map distance between new markers mapping to the ends of the LGs was comparatively high in this

study. This supports the hypothesis of high recombination rates in distal regions of pearl millet chromosome postulated by previous studies [2,16,45]. Gene-rich regions are usually subject to higher recombination rates compared to gene-free regions, which are mainly composed of repetitive DNA [46]. Four additional EST-SSR loci (*viz.*, *Xicmp3024*, *Xicmp3025*, *Xicmp3045* and *Xicmp3066*), which were monomorphic in the ICMB 841-P3 x 863B-P2 cross, were mapped using the F<sub>2</sub> population of the cross 81B-P8 x IPC 804 (data not shown). *Xicmp3024* was mapped to LG 6, *Xicmp3025* to LG 3, *Xicmp3045* to LG 2 and *Xicmp3066* to LG 7.

### Comparative Genomics

To confirm map locations of the new EST-SSR markers, the mapped EST microsatellite sequences were aligned against the rice genome. While six of the mapped pearl millet EST-SSR sequences found no hits on the rice genome, the remaining EST sequences had one or more rice homologues that were physically mapped. There was correspondence between pearl millet linkage groups and rice chromosomes as expected from rice-pearl millet synteny [45]. The homologous rice sequences for each pearl

millet EST sequence corresponding to a mapped SSR, and their putative functions, are given (see Additional file 4).

The integration of RFLP markers previously mapped in other grass species has provided the anchor points to align the pearl millet linkage groups to other cereal genetic maps. The complex relationship of pearl millet with foxtail millet and rice has been studied using RFLP probes [45]. So far most of the comparative genomics efforts have relied heavily on the hybridization-based RFLP technique with lower resolution of microsynteny [47]. However, recently, sequence-based comparative maps have been developed for rice-wheat [48] and sorghum-rice [49] that enhance the resolution considerably. The application of a PCR-based co-dominant marker system for comparative genomics would be highly desirable, because such a marker system can increase the efficiency of transferring genetic information across species. Due to their transferability, EST-SSR markers have good potential for such applications in comparative genomics [31,36,50]. Thus the EST-SSR developed in this study offer great potential in comparative genomics studies in pearl millet with other cereals.

## Conclusion

The frequency of SSR-containing ESTs was quite high in this initial sample of the expressed portion of the pearl millet genome, so it appears that this approach can add reasonable numbers of SSRs to the existing pearl millet SSR collection at very modest cost provided that 1) the sequence information is available freely (or at very low cost) as a result of other research programs, 2) care is taken to minimize redundancy, and 3) primer synthesis and testing is limited to only sequences flanking the most highly repeated di-, tri-, and tetra-nucleotide motifs. This study has demonstrated the potential utility of EST-derived SSR primers in pearl millet. As reported for other crops, EST-derived SSRs provide a cost-saving marker development option in pearl millet. These resources will add a sizeable number of relatively more useful SSRs to the existing repertoire of circa 100 genomic SSRs that are already available to pearl millet researchers. These new SSR markers are being used in our on-going marker-aided backcross program for improvement of downy mildew resistance, drought tolerance and stover yield and quality traits in pearl millet.

## Methods

### Identification of SSR and primer designing

In this study, 3520 pearl millet EST sequences (1.46 Mb) obtained from cDNA libraries constructed from root and shoot tissues of seedlings exposed to cold, drought and salinity stresses were used. These EST sequences were made available by our collaborators in International Center for Genetic Engineering and Biotechnology

(ICGEB), New Delhi and University of Hyderabad, Hyderabad, India. Out of 3520 sequences, 2810 are currently available in the National Center for Biotechnology Information (NCBI) GenBank database (accession numbers [32275159-32277652](#) and [91982272-91982588](#)). These EST sequences were assembled using CAP3 [51] to identify unigenes. The unigene sequences (singlets and contigs) were then searched for the presence of SSR motifs using the Micro Satellite identification tool (MISA) available at <http://pgrc.ipk-gatersleben.de/misa/misa.html>. The search was restricted to motifs having at least 14 bp length (*i.e.*, di-nucleotide > 7; tri-nucleotide > 5; tetra-nucleotide > 4; penta-nucleotide > 3 and hexa-nucleotide > 3) and the minimum difference between two SSRs in a sequence (interruption) was 100 bases. From the SSR-containing EST sequences, primer pairs were designed targeting the SSRs except hexa-nucleotide repeats using Primer3 [52]. The primer designing conditions were: 57–61 °C melting temperature, 40–60% GC content, and 18–24 bp primer length.

### Plant materials

The primer pairs developed from SSR-containing ESTs (prefixed with 'ICMP' for ICRISAT Millet Primers), hereafter referred to as 'EST-SSRs' along with a set of 18 SSR primer pairs (prefixed with 'CTM'), hereafter referred to as 'genomic SSRs' developed from small insert genomic library of pearl millet [18] were tested for their ability to detect polymorphism among 11 pairs of pearl millet mapping population parental lines available at ICRISAT-Patancheru *viz.*, PRLT 2/89-33 and H 77/833-2; ICMB 841-P3 and 863B-P2; Tift 23D<sub>2</sub>B<sub>1</sub>-P5 and WSIL-P8; PT 732B-P2 and P 1449-P1; LGD 1-B-10 and ICMP 85410-P7; 81B-P6 and ICMP 451-P8; ICMP 451-P6 and H 77/833-2-P5(NT); W 504-1-P1 and P 310-17; IP 18293-P152 and Tift 238D<sub>1</sub>-P158; ICMB 89111-P6 and ICMB 90111-P6; 81B-P8 and IPC 804. The origin and characteristics of these 22 inbred lines are given (see Additional file 5).

Two mapping populations were used to map the polymorphic markers. The first population consists of 149 F<sub>2</sub> individuals of the cross ICMB 841-P3 × 863B-P2, which has been previously used for mapping drought tolerance, downy mildew resistance, and stover quality QTLs [10,11,13]. The other population consists of 397 F<sub>2</sub> individuals of the cross 81B-P8 × IPC 804, which was originally developed to map fertility restorer genes [53].

### DNA extraction

Genomic DNA was extracted from leaf tissue of F<sub>3</sub> bulks representing each of the individual entries from the segregating F<sub>2</sub> population as described by Sharp *et al.* [54]. The DNA concentrations were determined by comparing the sample intensity with that of known amounts of uncut lambda DNA by electrophoresis in 1.2% agarose gels con-

taining ethidium bromide. DNA was diluted to a working stock concentration of approximately 5 ng/μl and checked before use.

### PCR and Electrophoresis

PCRs were carried out in a 10 μl reaction mixture containing 10–15 g of genomic DNA, 2 pmol of each primer, 1 mM MgCl<sub>2</sub>, 0.1 mM of each dNTP, 1× reaction buffer, and 0.2 U *Taq* polymerase (Bioline). After one denaturing step of 3 min at 94°C, a touchdown amplification program was performed on GeneAmp 9700 thermal cycler (Applied Biosystems, USA). This profile consisted of a denaturing step of 25 sec at 94°C and an extension step of 30 sec at 72°C. The initial annealing step was 20 sec at 64°C for one cycle and subsequently the temperature was reduced by 1°C for every cycle until a final temperature of 55°C was reached. The annealing temperature of 55°C was maintained for the last 35 cycles of amplification, followed by final extension of 72°C for 7 min. PCR products were size-separated on native polyacrylamide gels (6%) run on 0.5× TBE buffer at 600 V for 3 hours using a BioRad® sequencing gel apparatus. After electrophoresis, the banding patterns of PCR product on PAGE gels were visualized by silver-staining [55].

### Polymorphism assessment and mapping of EST-SSRs

The new markers were placed onto the already available framework linkage maps of the crosses ICMB 841-P3 × 863B-P2 with 55 RFLP and 32 SSR markers [10,11] and 81B-P8 × IPC 804 with 11 RFLP, 23 SSR and 3 morphological markers (53), using Mapmaker 3.0 [56] at LOD 4. The 'build' command with the LOD of 4 was used to incorporate SSR markers into the frameworks. Local orders were verified with the 'ripple' command (window size of 5 loci, LOD 3). Map distances were calculated based on the Haldane mapping function. The mapped EST-SSRs were aligned against the rice genome using the BLAST (WU-BLAST 2.0) against the IRGSP pseudomolecules Build04 <http://rgp.dna.affrc.go.jp/E/IRGSP/Build4/build4.html>. Only those sequences from rice that returned an e-value of e-10 or less during the BLAST search were considered putative homologs. BLAST searches were carried out in June 2006 and again validated in April 2008.

### Authors' contributions

SS designed the study, analysed the data and drafted the manuscript. BJ and VM were involved with *in-silico* analysis. PSK and SN performed initial optimization of primers and a part of wet lab experiments. TN was partly involved in data generation and analysis. CTH provided overall guidance in designing and performing this study and edited the manuscript. All the authors read and approved the manuscript.

## Additional material

### Additional file 1

*Details of newly synthesized EST-SSR primer pairs.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2229-8-119-S1.doc>]

### Additional file 2

*Polymorphism information of EST-SSRs on parental lines of pearl millet mapping populations.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2229-8-119-S2.doc>]

### Additional file 3

*Polymorphism information of genomic SSRs on parental lines of pearl millet mapping populations.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2229-8-119-S3.doc>]

### Additional file 4

*Positions of rice homologs of mapped ESTs and their putative functions.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2229-8-119-S4.doc>]

### Additional file 5

*Details of pearl millet inbred lines used for polymorphism testing.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2229-8-119-S5.doc>]

## Acknowledgements

The authors acknowledge Dr MK Reddy from the International Centre for Genetic Engineering and Biotechnology (ICGEB) and Dr AR Reddy from the Central University of Hyderabad for providing the EST sequences, and Mr M Narsi Reddy of ICRISAT for technical assistance throughout this study. This document is an output from a project (Plant Sciences Research Programme R8183) funded by the UK Department for International Development (DFID) and administered by CAZS Natural Resources for the benefit of developing countries. The views expressed are not necessarily those of DFID. Financial support from the Generation Challenge Programme (GCP) of the Consultative Group on International Agricultural Research (CGIAR) is also gratefully acknowledged.

## References

1. Gale MD, Devos KM, Zhu JH, Allouis S, Couchman MS, Liu H, Pittaway TS, Qi XQ, Kolesnikova-Allen M, Hash CT: **New molecular marker technologies for pearl millet improvement.** *J SAT Agric Res* 2005, **1**(1):.
2. Liu CJ, Witcombe JR, Pittaway TS, Nash M, Hash CT, Busso CS, Gale MD: **An RFLP-based genetic map of pearl millet (*Pennisetum glaucum*).** *Theor Appl Genet* 1994, **89**:481-487.
3. Qi X, Pittaway TS, Lindup S, Liu H, Waterman E, Padi FK, Hash CT, Zhu J, Gale MD, Devos KM: **An integrated genetic map and a new set of simple sequence repeat markers for pearl millet, *Pennisetum glaucum*.** *Theor Appl Genet* 2004, **109**:1485-1493.

4. Jones ES, Liu CJ, Gale MD, Hash CT, Witcombe JR: **Mapping quantitative trait loci for downy mildew resistance in pearl millet.** *Theor Appl Genet* 1995, **91**:448-456.
5. Jones ES, Breese WA, Liu CJ, Singh SD, Shaw DS, Witcombe JR: **Mapping quantitative trait loci for resistance to downy mildew in pearl millet: Field and glasshouse screens detect the same QTL.** *Crop Sci* 2002, **42**:1316-1323.
6. Morgan RN, Wilson JP, Hanna WW, Ozais-Akins P: **Molecular markers for rust and pyricularia leaf spot disease resistance in pearl millet.** *Theor Appl Genet* 1998, **96**:413-420.
7. Breese WA, Hash CT, Devos KM, Howarth CJ: **Pearl millet genomics – an overview with respect to breeding for resistance to downy mildew.** In *Sorghum and Millets Pathology 2000* Edited by: Leslie JF. Ames, Iowa, Iowa State Press; 2002:243-246.
8. Gulia SK, Hash CT, Thakur RP, Breese WA, Sangwan RS: **Mapping new QTLs for improvement of downy mildew resistance in pearl millet.** In *Crop Production in Stress Environments: Genetic and Management Options* Edited by: Singh DP, Tomar VS, Behl RK, Upadhyaya SD, Bhale MS, Khare B. Jodhpur, Rajasthan, India, Agrobios International; 2007:373-386.
9. Yadav RS, Hash CT, Bidinger FR, Cavan GP, Howarth CJ: **Quantitative trait loci associated with traits determining grain and stover yield in pearl millet under terminal drought-stress conditions.** *Theor Appl Genet* 2002, **104**:67-83.
10. Yadav RS, Hash CT, Bidinger FR, Devos KM, Howarth CJ: **Genomic regions associated with grain yield and aspects of post-flowering drought tolerance in pearl millet across stress environments and tester background.** *Euphytica* 2004, **136**:265-277.
11. Bidinger FR, Nepolean T, Hash CT, Yadav RS, Howarth CJ: **Quantitative trait loci for grain yield in pearl millet under variable postflowering moisture conditions.** *Crop Sci* 2007, **47**:969-980.
12. Yadav RS, Bidinger FR, Hash CT, Yadav YP, Yadav OP, Bhatnagar SK, Howarth CJ: **Mapping and characterization of QTL × E interactions for traits determining grain and stover yield in pearl millet.** *Theor Appl Genet* 2003, **106**:512-520.
13. Hash CT, Bhasker Raj AG, Lindup S, Sharma A, Beniwal CR, Folkertsma RT, Mahalakshmi V, Zerbini E, Blümmel M: **Opportunities for marker-assisted selection (MAS) to improve the feed quality of crop residues in pearl millet and sorghum.** *Field Crop Res* 2003, **84**:79-88.
14. Hash CT, Sharma A, Kolesnikova-Allen MA, Singh SD, Thakur RP, Bhasker Raj AG, Ratnaji Rao MNV, Nijhawan DC, Beniwal CR, Sagar P, Yadav HP, Yadav YP, Srikant , Bhatnagar SK, Khairwal IS, Howarth CJ, Cavan GP, Gale MD, Liu C, Devos KM, Breese WA, Witcombe JR: **Teamwork delivers biotechnology products to Indian small-holder crop-livestock producers: Pearl millet hybrid "HHB 67 Improved" enters seed delivery pipeline.** *J SAT Agric Res* 2006, **2**(1): [<http://www.icrisat.org/journal/bioinformatics/v2i1/teamwork.pdf>].
15. Khairwal IS, Hash CT: **"HHB 67-Improved" – The first product of marker-assisted crop breeding in India.** *Asia-Pacific Consortium on Agricultural Biotechnology (APCoAB) e-News* 2007 [[http://www.apcoab.org/special\\_news.html](http://www.apcoab.org/special_news.html)].
16. Qi X, Lindup S, Pittaway TS, Allouis S, Gale MD, Devos KM: **Development of simple sequence repeat markers from bacterial artificial chromosomes without subcloning.** *Biotechniques* 2001, **31**(2):355, 358-362.
17. Allouis S, Qi X, Lindup S, Gale MD, Devos KM: **Construction of a BAC library of pearl millet, *Pennisetum glaucum*.** *Theor Appl Genet* 2001, **102**:1200-1205.
18. Budak H, Pedraza F, Cregan PB, Baenziger PS, Dweikat : **Development and utilization of SSRs to estimate the degree of genetic relationships in a collection of pearl millet germplasm.** *Crop Sci* 2003, **43**:2284-2290.
19. Mariac C, Luong V, Kapran I, Mamadou A, Sagnard F, Deu M, Chantereau J, Gerard B, Ndjeunga J, Bezançon G, Pham J-L, Vigouroux Y: **Diversity of wild and cultivated pearl millet accessions (*Pennisetum glaucum* [L.] R. Br.) in Niger assessed by microsatellite markers.** *Theor Appl Genet* 2006, **114**:49-58.
20. Yadav OP, Mitchell SE, Zamora A, Fulton TM, Kresovich S: **Development of new simple sequence repeat markers for pearl millet.** *J SAT Agr Res* 2007, **3**(1): [[http://www.icrisat.org/journal/mpiv3i1/Sorghu Millet%20 other Cereals/ ISMN%2022\\_Development%20of%20new...%20\(OPYa dav%20et%20al\).pdf](http://www.icrisat.org/journal/mpiv3i1/Sorghu%20Millet%20other%20Cereals/ISMN%2022_Development%20of%20new...%20(OPYa%20dav%20et%20al).pdf)].
21. Morgante M, Hanafey M, Powell W: **Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes.** *Nat Genet* 2002, **30**:194-200.
22. Li YC, Korol AB, Fahima T, Nevo E: **Microsatellites within genes: structure, function and evolution.** *Mol Bio Evol* 2004, **21**:991-1007.
23. Miyao A, Zhong HS, Monna L, Yano M, Yamamoto K, Havukkala I, Minobe Y, Sasaki T: **Characterization and genetic mapping of simple sequence repeats in the rice genome.** *DNA Res* 1996, **3**:233-238.
24. Varshney RK, Graner A, Sorrells ME: **Genic microsatellite markers in plants: features and applications.** *Trends Biotechnol* 2005, **23**:48-55.
25. Nicot N, Chiquet V, Gandon B, Amilhat L, Legeai F, Leroy P, Bernard M, Sourdille P: **Study of simple sequence repeat (SSR) markers from wheat expressed sequence tags (ESTs).** *Theor Appl Genet* 2004, **109**:800-805.
26. Holton TA, Christopher JT, McClure L, Harker N, Henry RJ: **Identification and mapping of polymorphic SSR markers from expressed gene sequences of barley and wheat.** *Mol Breeding* 2002, **9**:63-71.
27. Wang ML, Barkley NA, Yu JK, Dean RE, Newman ML, Sorrells ME, Pederson GA: **Transfer of simple sequence repeat (SSR) markers from major cereal crops to minor grass species for germplasm characterization and evaluation.** *Plant Genetic Resources* 2005, **3**:45-57.
28. Pflieger S, Lefebvre V, Causse M: **The candidate gene approach in plant genetics: A review.** *Mol Breeding* 2001, **7**:275-291.
29. Chee PW, Rong J, Williams-Coplin D, Schulze SR, Paterson AH: **EST derived PCR-based markers for functional gene homologues in cotton.** *Genome* 2004, **47**:449-462.
30. Varshney RK, Thiel T, Stein N, Langridge P, Graner A: **In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species.** *Cell Mol Biol Lett* 2002, **7**:537-546.
31. Kantety RV, Rotal ML, Matthews DE, Sorrells ME: **Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat.** *Plant Mol Biol* 2002, **48**:501-510.
32. Jayashree B, Punna R, Prasad P, Bantte K, Hash CT, Chandra S, Hoisington DA, Varshney RK: **A database of simple sequence repeats from cereal and legume expressed sequence tags mined in silico: survey and evaluation.** In *Silico Biol* 2006, **6**:0054.
33. Morgante M, Olivieri AM: **PCR-amplified microsatellites as markers in plant genetics.** *Plant J* 1993, **3**:175-182.
34. Powell W, Machray GC, Provan J: **Polymorphism revealed by simple sequence repeats.** *Trends Plant Sci* 1996, **1**:215-222.
35. Temnykh S, Declerk G, Lukashover A, Lipovich L, Cartinhour S, McCouch S: **Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length-variation, transposon associations and genetic marker potential.** *Genome Res* 2001, **11**:1441-1452.
36. Thiel T, Michalek W, Varshney RK, Graner A: **Exploiting EST database for the development and characterization of gene derived SSR-markers in barley (*Hordeum vulgare* L.).** *Theor Appl Genet* 2003, **106**:411-422.
37. Gupta PK, Rustgi S, Sharma S, Singh R, Kumar N, Balyan HS: **Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat.** *Mol Genet Genomics* 2003, **270**:315-323.
38. Yu J-K, La Rota M, Kantety RV, Sorrells ME: **EST derived SSR markers for comparative mapping in wheat and rice.** *Mol Genet Genomics* 2004, **271**:742-751.
39. Saha MC, Mian MAR, Eujayl I, Zwonitzer JC, Wang L, May D: **Tall fescue EST-SSR markers with transferability cross several grass species.** *Theor Appl Genet* 2004, **109**:783-791.
40. Varshney RK, Grosse I, Hahnel U, Siefken R, Prasad M, Stein N, Langridge P, Altschmied L, Graner A: **Genetic mapping and BAC assignment of EST-derived SSR markers shows non-uniform distribution of genes in the barley genome.** *Theor Appl Genet* 2006, **113**:239-250.
41. Sharopova N, McMullen MD, Schultz I, Schroeder S, Sanchez-Villeda H, Gardiner , Bergstrom D, Houchins K, Melia-Hancock S, Musket T, Duru N, Polacco M, Edwards K, Ruff T, Register JC, Brouwer C, Thompson R, Velasco R, Chin E, Lee M, Woodman-Clíkeman W, Long MJ, Liscum E, Cone K, Davis G, Coe EH: **Development and**



- mapping of SSR markers for maize. *Plant Mol Biol* 2002, **48**:463-481.
42. Anderson JA, Ogihara Y, Sorrells ME, Tanksley SD: **Development of a chromosomal arm map for wheat based on RFLP markers.** *Theor Appl Genet* 1992, **83**:1035-1043.
  43. Cordeiro GM, Casu R, McIntyre CL, Manners JM, Henry RJ: **Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to erianthus and sorghum.** *Plant Sci* 2001, **160**:1115-1123.
  44. Gutierrez MV, Vaz Patto MC, Hugué T, Cubero JJ, Moreno MT, Torres AM: **Cross-species amplification of *Medicago truncatula* microsatellites across three major pulse crops.** *Theor Appl Genet* 2005, **110**:1210-1217.
  45. Devos KM, Pittaway TS, Reynolds A, Gale MD: **Comparative mapping reveals a complex relationship between the pearl millet genome and those of foxtail millet and rice.** *Theor Appl Genet* 2000, **100**:190-198.
  46. Wicker T, Zimmermann W, Perovic D, Paterson AH, Ganai M, Graner A, Stein N: **A detailed look at 7 million years of genome evolution in a 439 kb contiguous sequence at the barley *Hv-eif4e* locus: recombination, re-arrangements, and repeats.** *Plant J* 2005, **41**:184-194.
  47. Kilian A, Chen J, Han F, Steffenson B, Kleinhofs A: **Towards map-based cloning of the barley stem rust resistance genes *Rpg1* and *rpg4* using rice as an intergenomic cloning vehicle.** *Plant Mol Biol* 1997, **35**:187-195.
  48. Sorrells ME, La Rota M, Bermudez-Kandianis CE, Greene RA, Kantety R, Munkvold JD, Miftahudin M, Mahmoud A, Ma X, Gustafson PJ, Qi LL, Echalié B, Gill BS, Matthews DE, Lazo GR, Chao S, Anderson OD, Edwards H, Linkiewicz AM, Dubcovsky J, Akhunov ED, Dvorak J, Zhang D, Nguyen HT, Peng J, Lapitan NLV, Gonzalez-Hernandez JL, Anderson JA, Hossain K, Kalavacharla V, Kianian SF, Choi DW, Close TJ, Dilbirli M, Gill KS, Steber C, Walker-Simmons MK, McGuire PE, Qualset CO: **Comparative DNA sequence analysis of wheat and rice genomes.** *Genet Res* 2003, **13**:1818-1827.
  49. Klein PE, Klein RR, Vrebalov J, Mullet JE: **Sequence based alignment of sorghum chromosome 3 and rice chromosome 1 reveals extensive conservation of gene order and one major chromosomal rearrangement.** *Plant J* 2003, **34**:605-621.
  50. Eujayl I, Sledge MK, Wang L, May GD, Chekhovskiy K, Zwonitzer JC, Mian MAR: ***Medicago truncatula* EST-SSRs reveal cross-species genetic markers for *Medicago* spp.** *Theor Appl Genet* 2004, **108**:414-422.
  51. Huang X, Madan A: **CAP3: A DNA sequence assembly program.** *Genome Res* 1999, **9**:868-877.
  52. Rozen S, Skaletsky HJ: **Primer3 on the WWW for general users and for biologist programmers.** In *Methods Mol Biol Volume 132*. Edited by: Krawetz S, Misener S. Totowa, NJ, Humana Press; 2000:365-386.
  53. Yadav D: **Genetics of cytoplasmic-nuclear male sterility and identification of molecular markers of fertility restoration in pearl millet (*Pennisetum glaucum* (L.) R. Br).** In *PhD thesis* Chaudhary Charan Singh Haryana Agricultural University, Hisar, Haryana, India; 2005.
  54. Sharp PJ, Kries M, Sherry PR, Gale MD: **Location of  $\beta$ -amylase sequences in wheat and its relatives.** *Theor Appl Genet* 1988, **75**:286-290.
  55. Tegelstrom H: **Detection of mitochondrial DNA fragments.** In *Molecular Genetic Analysis of Populations: A Practical Approach* Edited by: Hoelzel R. Oxford, UK, IRL Press; 1992:89-114.
  56. Lander E, Green P, Abrahamson J, Barlow A, Daley M, Lincoln S, Newburg L: **MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations.** *Genomic* 1987, **1**:174-181.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

