Theor Appl Genet (2007) 114:321–332
DOI 10.1007/s00122-006-0435-7

ORIGINAL PAPER

The genetic map of finger millet, *Eleusine coracana*

Mathews M. Dida · Srinivasachary · Sujatha Ramakrishnan · Jeffrey L. Bennetzen · Mike D. Gale · Katrien M. Devos

Received: 23 April 2006 / Accepted: 12 October 2006 / Published online: 14 November 2006

Abstract Restriction fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP), expressed-sequenced tag (EST), and simple sequence repeat (SSR) markers were used to generate a genetic map of the tetraploid finger millet (*Eleusine coracana* subsp. *coracana*) genome ($2n = 4x = 36$). Because levels of variation in finger millet are low, the map was generated in an inter-subspecific F_2 population from a cross between *E. coracana* subsp. *coracana* cv. Okhale-1 and its wild progenitor *E. coracana* subsp. *africana* acc. MD-20. Duplicated loci were used to identify homoeologous groups. Assignment of linkage groups to the A and B genome was done by comparing the hybridization patterns of probes in Okhale-1, MD-20, and *Eleusine indica* acc. MD-36. *E. indica* is the A genome donor to *E. coracana*. The maps span 721 cM on the A genome and 787 cM on the B genome and cover all 18 finger millet chromosomes, at least partially. To facilitate the use of marker-assisted selection in finger millet, a first set of 82 SSR markers was developed. The SSRs were identified in small-insert genomic libraries generated using methylation-sensitive restriction enzymes. Thirty-one of the SSRs were mapped. Application of the maps and markers in hybridization-based breeding programs will expedite the improvement of finger millet.

Electronic supplementary material Supplementary material is available in the online version of this article at http://dx.doi.org/10.1007/s00122-006-0435-7 and is accessible for authorized users.

Communicated by E. Guiderdoni.

M. M. Dida · Srinivasachary · M. D. Gale · K. M. Devos
John Innes Centre, Norwich Research Park, Colney, Norwich, NR4 7UH, UK

M. M. Dida · Srinivasachary · K. M. Devos (✉)
Department of Crop and Soil Sciences and Department of Plant Biology, University of Georgia, Athens, GA 30602, USA
e-mail: kdevos@uga.edu

M. M. Dida
Department of Botany and Horticulture, Maseno University, P.O. Box 333, Maseno, Kenya

S. Ramakrishnan · J. L. Bennetzen
Department of Biological Sciences, Purdue University, West Lafayette, IN 47907, USA

J. L. Bennetzen
Department of Genetics, University of Georgia, Athens, GA 30602, USA

Introduction

Finger millet, *Eleusine coracana* (L.) Gaertn. subsp. *coracana*, is an important cereal in East Africa and Southern India. The crop is adapted to a wide range of environments, can withstand significant levels of salinity, is relatively resistant to water logging, and has few serious diseases. Finger millet is grown mainly by subsistence farmers and serves as a food security crop because of its high-nutritional value and excellent storage qualities. Under irrigated conditions in field trials, yields of up to 5–6 metric tonnes/ha have been obtained (National Research Council 1996). However, yields in farmers' fields, usually sown with unimproved varieties, are commonly between 400 and 2,000 kg/ha.

Breeding efforts in finger millet have been very limited. Crossing has long been impeded by the highly self-pollinating nature of the crop and the small flower

size that has made hand-emasculation difficult. The development in the 1960s of hot water emasculation, combined with contact pollination, opened the way for the development of new crossbred varieties. However, even today, all released varieties in Africa are germplasm selections. In India, the hybridization of Indian and African varieties has resulted in the production and release of high-yielding 'Indaf' types that have proven popular with farmers.

Finger millet was domesticated ~5,000 years ago from its wild progenitor, *E. coracana* subsp. *africana*. The A genome donor is believed to be *Eleusine indica*, a ubiquitous weed of tropical and subtropical regions, but there are conflicting reports on the B genome donor. Both *Eleusine tristachya* and *Eleusine floccifolia* have been considered potential B genome donors to *E. coracana* based on rDNA restriction patterns (Hilu and Johnson 1992) and genomic in situ hybridization (GISH) (Bisht and Mukai 2001), respectively. The latter study demonstrated that *E. tristachya* DNA hybridized to the same subset of *E. coracana* chromosomes as did *E. indica*, suggesting that *E. tristachya* is an A genome species. GISH patterns of *E. floccifolia* and *E. indica*, on the other hand, did not overlap. Subspecies *coracana* and *africana* hybridize readily, with more than 80% of the meiotic cells in the hybrid forming bivalents, suggesting that the two genomes are highly homoeologous (Hiremath and Salimath 1992).

Genetic knowledge and resources are scarce in finger millet. Finger millet is a tetraploid with genome composition AABB and a basic chromosome number of 9 ($2n = 4x = 36$). Its genome size (1C value) has been estimated by flow cytometry to be about 1.8 pg (Mysore and Baird 1997). This estimate is some 30% lower than the earlier measurements carried out by Feulgen microdensitometry (Hiremath and Salimath 1991). Nevertheless, the finger millet genome is still considered relatively large from a genetics/genomics perspective. Genetic research in finger millet has been limited to studying the mode of inheritance of a few qualitative traits reviewed by Rachie and Peters (1977) and biodiversity analyses. Isozyme and DNA marker analyses have indicated that cultivated finger millet has a narrow genetic base and most likely went through a bottleneck during domestication (Hilu and Johnson 1992; Werth et al. 1994; Muza et al. 1995; Salimath et al. 1995). As expected, variation in the wild subsp. *africana* was considerably higher (Hilu and Johnson 1992; Werth et al. 1994). Although the finger millet germplasm pool remains largely uncharacterized, small-scale analyses of the nutritional value of seeds of wild and cultivated *E. coracana* lines have shown a wide variation in protein, calcium, and iron content (Barbeau and Hilu 1993; Vadivoo et al. 1998). Phenotypic variation for blast resistance, early vigor and other yield-related characters has also been observed. Finger millet, thus, has great potential for improvement through the application of hybridization breeding. The construction of a finger millet genetic map, reported in this paper, is a first step toward mapping traits of agronomic importance and will ultimately help in trait transfer in breeding programs.

Materials and methods

Plant material

To aid in the choice of mapping parents, variation within and between *E. coracana* subspecies was assessed using four cultivars belonging to subsp. *coracana* and four wild accessions belonging to subsp. *africana*. The accessions, their region/country of origin and the seed sources are given in Table 1. In addition, two accessions of *E. indica*, the A genome donor to *E. coracana*, were also analyzed (Table 1). Accessions IE 1012 and PI 321125, which were used as sources for the development of expressed sequenced tags (ESTs) and simple sequence repeats (SSRs), respectively, were obtained from the Plant Genetic Resources Conservation Unit (PGRCU) of the USDA-ARS, GA, USA.

A mapping population consisting of 151 F_2 progeny was generated from a cross between *E. coracana* subsp.

Table 1 The origin and source of *Eleusine* accessions and cultivars used in the polymorphism screen and mapping analysis

Accession or cultivar	Origin	Source
subsp. *coracana*		
KNE 100007	Ethiopia	ICRISAT, Nairobi
Gulu-E	Uganda	KARI
Okhale-1	Nepal	KARI
Nanjala	Kenya	KARI
subsp. *africana*		
GBK 030647	Elgeyo Marakwet, Kenya	GBK, Muguga
MD-14	Bukura, Kenya	Collected
MD-20	Ragengni, Bondo, Kenya	Collected
MD-51	Kitale, Kenya	Collected
Eleusine indica		
MD-36	Kinango, Kwale	Collected
GBK 05528	Marsabit, Kenya	GBK, Muguga

ICRISAT International Crops Research Institute for the Semi-arid Tropics

KARI Kenya Agricultural Research Institute

GBK Gene Bank of Kenya

africana acc. MD-20 and *E. coracana* subsp. *coracana* cv. Okhale-1. Subsp. *africana* is the wild progenitor of cultivated finger millet, subsp. *coracana*. Okhale-1 is an elite cultivar from Nepal, and was obtained from the Kenya Agricultural Research Institute (KARI). MD-20 was collected in the wild at Ragengni, in the Bondo district of Kenya.

Markers

RFLPs

*Pst*I genomic probes (*Xpse* prefix) in the size range of 0.5–2.5 kb were generated by digesting DNA from *E. coracana* subsp. *coracana* acc. MD-9 (collected from Vihiga, Kenya) with the restriction enzyme *Pst*I, purifying the fragments in the selected size range and cloning them into a pBluescript (Stratagene, La Jolla, CA, USA) vector. Heterologous probes were obtained from the Rice Genome Program, Tsukuba, Japan (*Xrgc, Xrgr, Xrgl* prefixes for callus and root ESTs, and *Not*I linking genomic clones, respectively), Cornell University (oat cDNAs, *Xcdo* prefix) and the John Innes Centre (wheat cDNA and *Pst*I genomic clones, *Xpsr* prefix). Two disease resistance gene analogs, one isolated from finger millet (*Xrga6_10*) and one from pearl millet (*Xrga216*), were obtained from Dr. S. Sivaramakrishnan, Acharya NG Ranga Agricultural University, Hyderabad, India. Methods for DNA isolation, restriction enzyme digestion, gel electrophoresis, Southern transfer, probe labeling and hybridization were as described in Devos et al. (1992). Two restriction enzymes were used; *Eco*RI and *Hin*dIII.

AFLPs

The amplified fragment length polymorphism (AFLP) protocol used was essentially as described in Vos et al. (1995). *Mse*I and *Pst*I were used as frequent and rare cutting enzymes, respectively. One *Pst*I primer with three selective bases, GTT, was used in combination with one of four *Mse*I primers with selective bases G, AC, CA, and CG.

ESTs

Finger millet ESTs were obtained by sequencing cDNA clones derived from mRNA isolated from young leaves from subsp. *africana* acc. MD-20 (*Xlfm)* and subsp. *coracana* cv. Okhale-1 (*Xlfo*), roots from MD-20 seedlings (*Xrtm*), and immature inflorescences of subsp. *coracana* acc. IE 1012 (*Xinf*). The EST sequences can be obtained from GenBank (Okhale leaf ESTs: DY625711–DY625766, DY632421 and EB684283–EB684316; MD-20 leaf ESTs: DY684794–DY684907; MD-20 root ESTs: EB086105–EB086294; IE 1012 inflorescence ESTs: EB187296–EB187503). For finger millet ESTs for which an ortholog could be identified in the rice genomic sequence (Srinivasachary and K. M. Devos, unpublished data), the alignment between the finger millet and rice sequences was used to establish the positions of introns. Primers were designed using the Primer3 software (http://www.frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) and were preferentially made to span introns. PCR was carried out in 20 µl reaction volumes containing 1× PCR buffer, 1.25 mM $MgCl_2$, 100 ng finger millet template DNA, 500 nM each of the forward and reverse primer, 200 µM dNTPs and 0.8 U of Taq DNA polymerase (Promega, Madison, WI, USA). Amplification was performed using an initial denaturation step of 94°C for 3 min followed by 30 cycles of denaturation at 94°C for 50 s, annealing at 60°C for 50 s and 1.2 min extension at 72°C, and a final extension at 72°C for 4 min. Amplification products were separated by single strand conformation polymorphism in 0.5% Mutation Detection Enhancement (MDE) gels (Cambrex Bio Science, Walkersville, MD, USA) on standard 42 × 48 × 0.04 cm vertical electrophoresis systems. Gels were run overnight at room temperature at a constant power of 8 W after which DNA fragments were visualized by silver staining.

SSRs

Finger millet di- and trinucleotide SSRs (*Xugep*) were isolated from random genomic *Hin*dIII, *Pst*I and *Sal*I libraries of finger millet accession PI 321125 following hybridization of 18,432 double-gridded colonies with mixtures of $(AG)_{15}/(AC)_{15}$, $(AAG)_{10}/(AAC)_{10}/(ATC)_{10}$, $(AGC)_{10}/(AGG)_{10}$, and $(CCG)_{10}$ oligonucleotides. Three hundred and eighty-four positive clones were sequenced. Sequences of clones containing SSRs were used as queries in a BLASTx search against GenBank proteins, to establish whether the SSR was located in a 5′ untranslated region (UTR), 3′UTR, intron or coding region. The *E*-value cut-off for a positive hit was chosen arbitrarily as $1e^{-7}$.

Primers were designed against the SSR flanking sequences using the Primer3 software and tested for amplification using DNA from subsp. *africana* acc. MD-20 and subsp. *coracana* cv. Okhale-1 as template. PCR conditions were as described for ESTs. Amplification was carried out using a touchdown program consisting of an initial denaturation at 94°C for 3 min followed by 10 cycles of denaturation at 94°C for 30 s,

touchdown annealing starting at 62°C for 30 s and decreasing 0.7°C/cycle, and extension at 72°C for 1 min. This was followed by a further 35 cycles at an annealing temperature of 55°C. The program finished with a final extension at 72°C for 4 min. Amplification products were separated in 6% denaturing polyacrylamide gels and visualized by silver staining.

For primer pairs that showed variation between MD-20 and Okhale-1, an M13-tailed forward primer (M13-tail: 5′-CGTTGTAAAACGACGGCCAGT-3′) was synthesized. For mapping, SSR-specific M13-tailed forward and reverse primers were used in combination with a fluorescently labeled M13 primer to generate fluorescently labeled amplification products of a subset of 94 progeny that were separated on an ABI 3730×l sequencer. PCR was done in a reaction volume of 20 µl consisting of 1× PCR buffer, 1.5 mM $MgCl_2$, 50 ng finger millet template DNA, 100 nM each of the M13-tailed specific forward and labeled universal M13 primer, 160 nM of the specific reverse primer, 175 µM dNTPs and 0.7 U of AmpliTaq Gold (Applied Biosystems, Foster City, CA, USA) or Taq DNA polymerase (Promega). Thermocycling conditions were essentially as described above for SSRs, with the exception that the initial denaturation was done for 8 min, the touchdown was done in steps of 1°C and for 2 cycles at each temperature, and the total number of cycles at 55°C was reduced to 21. Samples were diluted either fivefold (samples labeled with PET) or tenfold (for 6-FAM, VIC and NED dyes) (Applied Biosystems). Four SSRs, each labeled with a different fluorochrome, were run together with a GS500LIZ-3730 standard (Applied Biosystems) on an ABI 3730×l. Patterns were analyzed using GeneMapper Version 3.5 (Applied Biosystems).

Mapping

Genetic maps were constructed using the program Mapmaker Version 3.0 (Whitehead Institute for Biomedical Research, Cambridge, MA, USA). Recombination frequencies were converted to centiMorgan distances using the Kosambi function. Initially, linkage groups were formed at a LOD cut-off of 11 and a maximum distance of 50 cM. This led to more linkage groups than the haploid chromosome number in finger millet. Linkage groups that were split at LOD 11, but joined at lower LOD scores, were then investigated using three-point and multipoint analyses for true linkage. Marker orders established using Mapmaker were verified manually by looking at the location of crossover events. Some markers could be placed in multiple locations with equal likelihood, and the range of locations is indicated on the map with a vertical bar (Fig. 1).

Results

Variation within *E. coracana*

Variation within subspecies *coracana* and *africana*, and between the two subspecies, was initially assessed using 20 single-copy or two-copy restriction fragment length polymorphism (RFLP) markers on DNA isolated from four cultivated and four wild accessions digested with either *Eco*RI or *Hin*dIII. Within the cultivated subsp. *coracana*, no variation was observed. Within the wild subsp. *africana*, 35% of probe/enzyme combinations revealed polymorphism in one or other of the two genomes, and the number of informative markers climbed to 75% when assessing variation between subsp. *coracana* and *africana*.

SSR markers

Following hybridization of 3,072 *Hin*dIII, 6,144 *Sal*I, and 9,216 *Pst*I clones with di- and trinucleotide-motifs, 384 positive clones were selected and sequenced. The number of clones screened, the number of positive clones identified for sequencing following hybridization and those containing SSR motifs with more than seven and ten repeat units for trinucleotide and dinucleotide repeats, respectively, are given in Table 2. In addition to the 96 SSRs identified by hybridization and sequence analysis (Table 2), three additional SSRs were identified during the sequence analysis of 96 randomly selected clones from each of the three libraries that was carried out to test the quality of the libraries.

Among the sequenced clones, 37 and 11 members belonging to long terminal repeat (LTR) and long interspersed nuclear element (LINE)-like retroelement families, respectively, were identified that contained a $(CAA)_n$ SSR. In the LTR retrotransposon family, the $(CAA)_n$ SSR was embedded in the repeat and this precluded the generation of locus-specific primers. In the LINE-like retroelement family, the $(CAA)_n$ repeat marked the boundary with unique DNA. Universal primers located in the repeat, combined with a specific primer in the unique DNA, generated locus-specific amplification products. However, many of the primer combinations assessed the differential presence of the LINE-like element in Okhale-1 and MD-20 rather than SSR variation. Two such markers, *Xugep43* and *Xugep41*, were mapped on linkage groups 1B and 6Aa, respectively. SSRs in clones

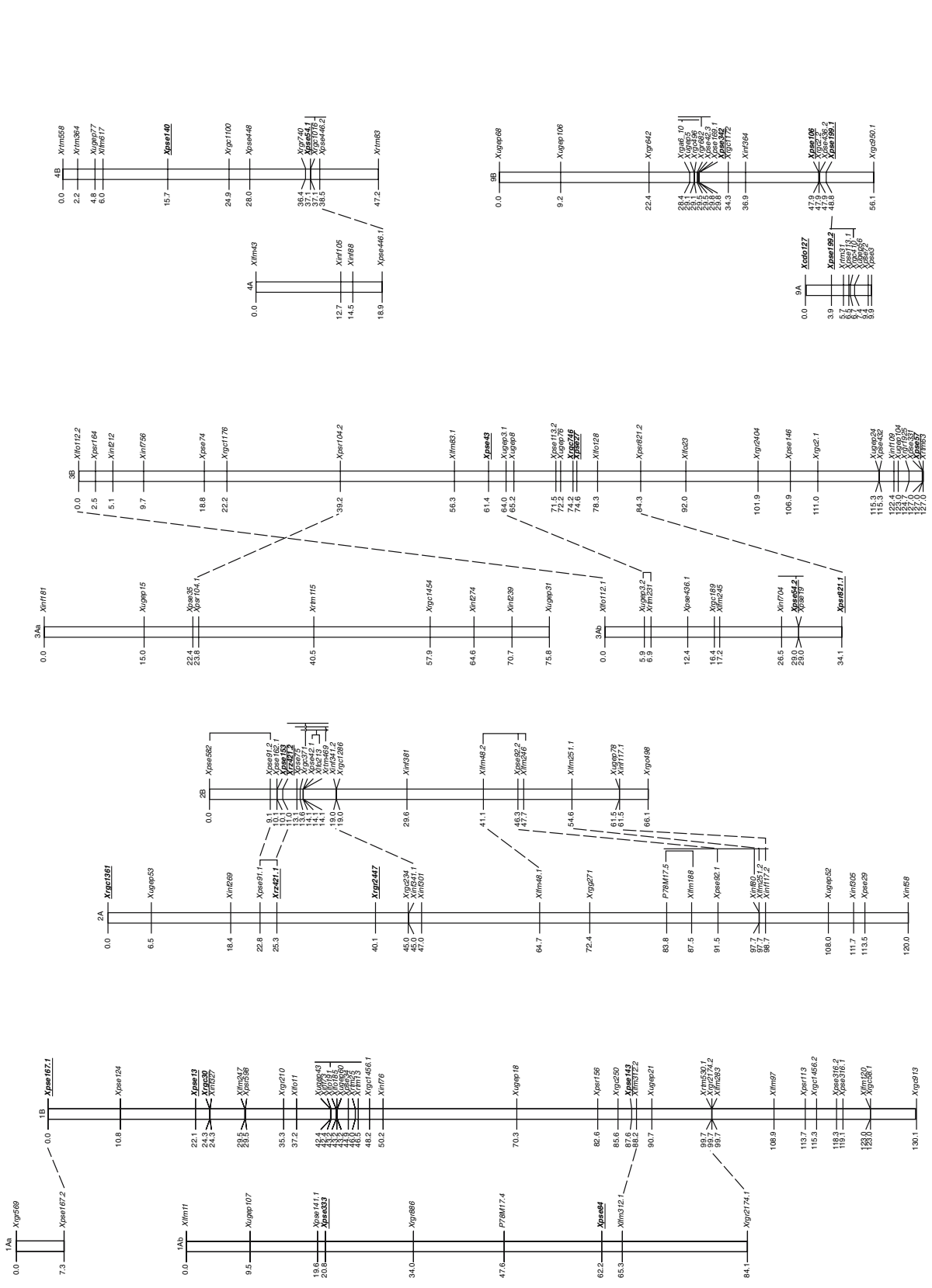

Fig. 1 Genetic maps of the finger millet genome. Loci are indicated on the *right-hand side* of the maps, genetic distances in cM (Kosambi) on the *left-hand side*. *Vertical bars* spanning some loci indicate ambiguous marker orders. Names of linkage groups are given above each map, with A or B indicating whether the linkage group originates from the A or B genome of finger millet. Loci in ***bold*** and *underlined* are the loci that were used to assign the linkage group to a genome

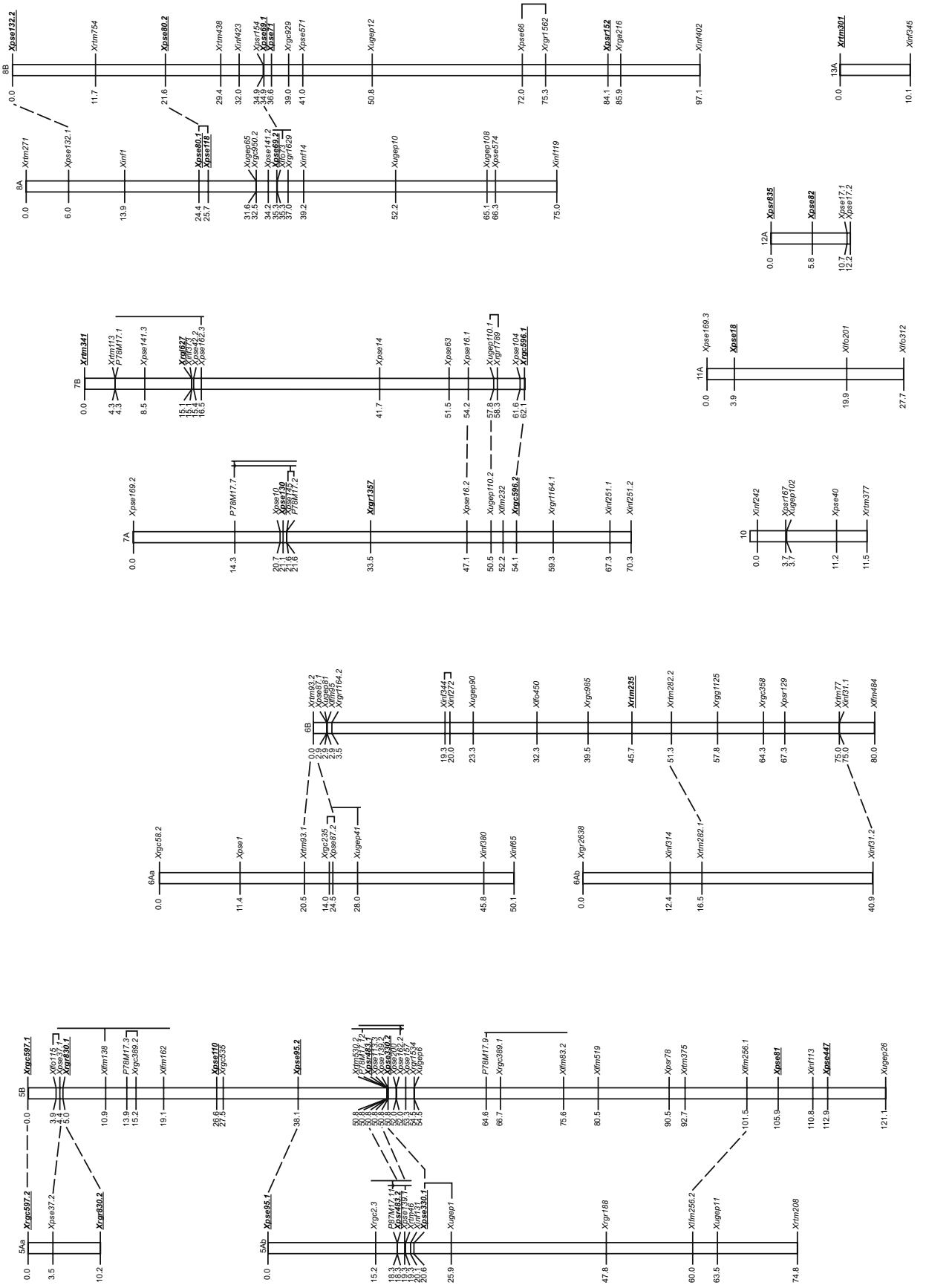

Fig. 1 continued

Table 2 Number of clones identified to carry SSR motifs

	Number of (%) positive clones identified by hybridization	Number of (%) SSRs identified by sequence analysis
*Hin*dIII—3,072 clones		
$(AG)_{15}/(AC)_{15}$	13 (0.42)	8 (0.26)
$(AAG)_{10}/(AAC)_{10}/(ATC)_{10}$	27 (0.88)	0
$(AGC)_{10}/(AGG)_{10}$	6 (0.20)	0
$(CCG)_{10}$	14 (0.46)	0
*Sal*I—6,144 clones		
$(AG)_{15}/(AC)_{15}$	79 (1.29)	28 (0.46)
$(AAG)_{10}/(AAC)_{10}/(ATC)_{10}$	14 (0.23)	2 (0.03)
$(AGC)_{10}/(AGG)_{10}$	29 (0.47)	4 (0.07)
$(CCG)_{10}$	36 (0.65)	1 (0.02)
*Pst*I—9,216 clones		
$(AG)_{15}/(AC)_{15}$	82 (0.89)	44 (0.48)
$(AAG)_{10}/(AAC)_{10}/(ATC)_{10}$	24 (0.26)	5 (0.05)
$(AGC)_{10}/(AGG)_{10}$	16 (0.18)	2 (0.02)
$(CCG)_{10}$	44 (0.49)	2 (0.02)

belonging to these repeat families were not included in Table 2.

A BLASTx search of the SSR-containing sequences against GenBank proteins allowed the location of 15 SSRs to either the untranslated or coding regions of genes. Five dinucleotide repeats were present in introns, two di- and one-trinucleotide repeat were in 5′UTRs, four dinucleotide repeats were in 3′UTRs, and three trinucleotide repeats were embedded in coding regions.

Primer pairs were designed against the flanking sequences of 99 SSRs and tested for their ability to amplify and detect polymorphism between the parents of the mapping population, Okhale-1 and MD-20. Eighty-two primer pairs (83%) gave clear amplification products, 62% of which displayed variation. Thirty-one (61%) of the polymorphic SSRs have been placed on the genetic map. The primer sequences and SSR motifs for the mapped and unmapped SSRs are given in Table 3 and Supplementary Table 1, respectively.

The genetic map

A total of 18 major linkage groups with seven or more markers were formed at LOD 11. A further ten minor linkage groups consisted of two to five markers. Initially, the maps were constructed using the LOD 11 grouping. Linkage groups and unlinked markers that were joined at lower LOD values were then merged and reanalyzed for true linkage by three point and multipoint analyses. Two of the major linkage groups were shown to be part of the same chromosome, and the number of minor groups was reduced from ten to eight. The maps are shown in Fig. 1. In total, 131 markers covering 721.4 cM on 16 linkage groups were mapped on the A genome, 196 markers covering 786.8 cM on nine linkage groups on the B genome, and a linkage group of five markers remained unassigned.

Assignment of LGs to homoeologous groups

Linkage groups that carried a minimum of two duplicated loci were assumed to be homoeologous. Based on this criterion, seven homoeologous groups were identified, five consisting of a pair of linkage groups, and two consisting of one linkage group showing homoeology with two linkage groups (Fig. 1). Eight further linkage groups carried between one and six multicopy loci, but did not show an obvious pattern of duplication. Six of these linkage groups could nevertheless be assigned to a homoeology group by a process of elimination. Of the 16 multicopy loci on unassigned linkage groups, five were triplicated and were not considered indicators for homoeology, two represented an intrachromosomal duplication and three duplicated loci fell into regions of already assigned homoeology. The remaining six loci, one per linkage group, were used to putatively assign these linkage groups to homoeology groups (Fig. 1). In total, nine homoeology groups were identified that correspond, presumably, to the nine sets of homoeologous *E. coracana* chromosomes.

Assignment of linkage groups within homoeologous groups to the A and B genomes

For most RFLP probe/enzyme combinations, the variation detected between subsp. *coracana* and *africana* was limited to only one of the genomes. Some probes also failed to detect variation between the A genomes

Table 3 Primer sequences, SSR repeat length, and PCR product length in acc. PI 321125 of the SSRs that were placed on the MD-20 × Okhale-1 genetic map

Primer	Primer sequence	SSR motif	Product length
UGEP1F	TTCAGTGGTGACGGAAGTTCT	$(TC)_{11}$	233
UGEP1R	GGCTCCATGAAGAGCTTGAC		
UGEP3F	CCACGAGGCCATACTGAATAG	$(CA)_7N_{12}(GA)_{15}$	206
UGEP3R	GATGGCCACTAGGGATGTTG		
UGEP5F	TGTACACAACACCACACTGATG	$(TC)_{12}AC(TC)_4$	215
UGEP5R	TTGTTTGGACGTTGGATGTG		
UGEP6F	AGCTGCAGTTTCAGTGGATTC	$(GA)_3TA(GA)_9$	229
UGEP6R	TCAACAAGGTGAAGCAGAGC		
UGEP8F	ATTTCCGCCATCACTCCAC	$(GA)_{13}$	297
UGEP8R	AGACGCAAATGGGTAAATGTC		
UGEP10F	AAACGCGATGAATTTTAAGCTC	$(GA)_{19}$	400
UGEP10R	CTATGTCGTGTCCCATGTCG		
UGEP11F	CCTCGAGTGGGGATCCAG	$(CT)_{12}$	153
UGEP11R	AAGACGCTGGTGGAAATAGC		
UGEP12F	ATCCCCACCTACGAGATGC	$(CT)_{22}$	230
UGEP12R	TCAAAGTGATGCGTCAGGTC		
UGEP15F	AAGGCAATCTCGAATGCAAC	$(CT)_{22}$	180
UGEP15R	AAGCCATGGATCCTTCCTTC		
UGEP18F	TTGCATGTGTTGCTTTTTGC	$(CT)_{12}$	318
UGEP18R	TGTTCTTGATTGCAAACTGATG		
UGEP21F	CAATTGATGTCATTGGGACAAC	$(GA)_{16}$	225
UGEP21R	GTATCCACCTGCATGCCAAC		
UGEP24F	GCCTTTTGATTGTTCAACTCG	$(GA)_{26}$	183
UGEP24R	CGTGATCCCTCTCCTCTCTG		
UGEP26F	ATGGGGTTAGGGTTCGAGTC	$(CGG)_7$	227
UGEP26R	TGTCCCTCACTCGTCTCCTC		
UGEP31F	ATGTTGATAGCCGGAAATGG	$(GA)_{12}$	241
UGEP31R	CCGTGAGCCTCGAGTTTTAG		
UGEP52F	TCATGCTAGCTTCAACACAACC	$(GA)_{16}$	215
UGEP52R	TGCTGGGTGAAACCCTAGAC		
UGEP53F	TGCCACAACTGTCAACAAAAG	$(AG)_{26}$	226
UGEP53R	CCTCGATGGCCATTATCAAG		
UGEP56F	CTCCGATACAGGCGTAAAGG	$(GT)_{12}$	162
UGEP56R	ACCATAATAGGGCCGCTTG		
UGEP60F	AGCTCTGCTTGGTGGAGAAG	$(GA)_{37}$	240
UGEP60R	TTTTCTACTGGTGGGCGAAG		
UGEP65F	AGTGCTAGCTTCCCATCAGC	$(CT)_{19}$	226
UGEP65R	ACCGAAACCCTTGTCAGTTC		
UGEP68F	CGGTCAGCATATAACGAATGG	$(CT)_{14}$	232
UGEP68R	TCATTGATGAATCCGACGTG		
UGEP76F	GCACGTACGGATTCACATTG	$(CAG)_7$	168
UGEP76R	GGTACGGAGACATCGACACC		
UGEP77F	TTCGCGCGAAATATAGGC	$(CT)_{19}$	245
UGEP77R	CTCGTAAGCACCCACCTTTC		
UGEP78F	AAGCAATCAACAAAGCCTTTTC	$(GA)_{14}$	244
UGEP78R	TACAACGTCCAGGCAACAAG		
UGEP81F	AAGGGCCATACCAACACTCC	$(GT)_{12}$	192
UGEP81R	CACTCGAGAACCGACCTTTG		
UGEP90F	GGCCTTTGCAGTCATGTGAG	$(CT)_{11}/(CT)_8$	232
UGEP90R	CGACTCCAGGTGTTGTTGG		
UGEP102F	ATGCAGCCTTTGTCATCTCC	$(TG)_{17}$	184
UGEP102R	GATGCCTTCCTTCCCTTCTC		
UGEP104F	TCAGCACCACCTGAATAGG	$(CT)_{11}$	189
UGEP104R	AATAGGGAGGGCGAAGACTC		
UGEP106F	AATTCCATTCTCTCGCATCG	$(AC)_{12}$	175
UGEP106R	TGCTGTGCTCCTCTGTTGAC		
UGEP107F	TCATGCTCCATGAAGAGTGTG	$(GA)_{15}$	224
UGEP107R	TGTCAAAAACCGGATCCAAG		
UGEP108F	GTTGGCTGCTCTGCTTATCC	$(CTG)_6(CAG)_2$	150
UGEP108R	TATCTGCTTGTGCAGCTTCG		
UGEP110F	AAATTCGCATCCTTGCTGAC	$(CT)_{12}$	192
UGEP110R	TGACAAGAGCACACCGACTC		

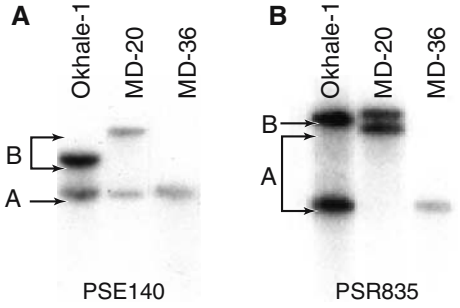

Fig. 2 Allocation of linkage groups to genomes. **a** Hybridization pattern obtained with PSE140 shows one fragment that is monomorphic in Okhale-1, MD-20, and MD-36 (*Eleusine indica*) and therefore located on the A genome. The fragment that is polymorphic between Okhale-1 and MD-20 and was mapped, assigns the locus *Xpse140* to the B genome. **b** Hybridization pattern of PSR835 shows a fragment that is monomorphic between MD-36 and Okhale-1, and thus expected to be derived from the A genome. This fragment is polymorphic between Okhale-1 and MD-20 and was mapped, allocating the locus *Xpsr835* to the A genome

of *E. coracana* and *E. indica*, the A genome donor, and this could be exploited to determine the genome origin of the linkage groups. The presence of a fragment of the same molecular weight in *E. indica*, *E. coracana* subsp. *Coracana*, and *E. coracana* subsp. *africana*, suggested that the monomorphic fragment belonged to the A genome and that the polymorphic fragment, the one that was mapped, belonged to the B genome (Fig. 2a). Other markers detected fragments that were monomorphic between *E. indica* and either *E. coracana* subsp. *coracana* or subsp. *africana* and polymorphic between the two subspecies. In this case, we could conclude that the A genome fragment had been mapped (Fig. 2b). The loci that were assigned to genomes are shown on the map (Fig. 1). The assignment of 17 linkage groups to either the A or B genome was supported by two or more loci and conforms to the expectation that a homoeology group consists of an A and B genome chromosome. Eight further linkage groups were allocated to either the A or B genome based on a single assigned locus or on the fact that the homoeologous linkage group had been assigned to a genome.

Discussion

Prevalence of SSRs in different fractions of the finger millet genome

The first generation of SSR markers in plants was developed from genomic DNA. In recent years, the large collections of ESTs that are available in some species have provided an alternative source of SSRs. EST-SSRs generally display lower levels of variation compared to genomic DNA, but tend to have higher transferability across species (La Rota et al. 2005 and references therein). The overall frequency of SSRs in plant genomes appears to be inversely correlated with genome size but is constant in the transcribed portion of the genome (Morgante et al. 2002; La Rota et al. 2005). Furthermore, Morgante and colleagues reported that, in all plant species examined, the SSR density was higher in the non-translated portions of transcribed regions than in the intergenic genomic DNA (Morgante et al. 2002). This suggests that the gene space is the best source for SSR markers.

In finger millet, however, the total number of ESTs available is less than 2000. Therefore, to generate SSRs from the gene space, libraries were constructed of finger millet genomic DNA digested with the methylation-sensitive enzymes *Sal*I (sensitive to CpG methylation) and *Pst*I (sensitive to CpNpG methylation). DNA was size selected in the range 1,300–1,700 bp, before cloning. It has been well established that transcribed regions are often hypomethylated and are thus preferentially present in fragments <3 kb generated by Cp(Np)G methylation-sensitive enzymes (Gruenbaum et al. 1981; Bennetzen et al. 1994). A third finger millet genomic library was made using the methylation-insensitive enzyme *Hin*dIII.

Screening of the three libraries with mixtures of di- and tri-nucleotide repeats showed that the overall frequency of SSRs was not very different in the three libraries (1.95, 2.57, and 1.80% for *Hin*dIII, *Sal*I, and *Pst*I, respectively). However, while the *Sal*I and *Pst*I libraries contained predominantly $(AG)_{15}/(AC)_{15}$ repeats, the *Hin*dIII library contained mainly $(AAG)_n/(AAC)_n/(ATC)_n$ trinucleotide repeats (Table 2). Sequence analysis showed that 88 and 12% of the clones that were positive upon hybridization with $(AG)_{15}/(AC)_{15}$ oligonucleotides were $(GA)_n$ and $(CA)_n$ SSRs, confirming the relative paucity of $(CA)_n$ SSRs in plants (Lagercrantz et al. 1993; Morgante et al. 2002). The difference in number of clones that contained SSRs based on hybridization assays and those identified as suitable for marker development based on sequence data can be explained by a number of factors. Some SSRs were shorter than the 21 and 20 bp threshold set for tri- and dinucleotide-repeats, respectively. Clones that contained multiple SSR motifs and duplicate clones were counted only once. As stated earlier, SSRs present in members of two retroelement families did not provide useful markers. Similarly, sequences that consisted of strings of SSRs and SSR-like repeats were considered unsuitable for primer design and were also discounted. Lastly, not all clones were fully

sequenced and SSRs may have been present in the unsequenced regions.

SSR sequences originating from retroelement families

All SSRs in the $(AAG)_n/(AAC)_n/(ATC)_n$ group in the *Hin*dIII library were $(CAA)_n$ repeats, and 93% (26 CAA SSRs representing 0.82% of the library) were present in members of the same LTR retrotransposon family. Elements of this LTR retrotransposon family were also present in the *Sal*I (3 CAA SSRs; 0.05% of the library) and *Pst*I libraries (8 CAA SSRs; 0.09% of the library). The remaining two $(CAA)_n$ SSRs in the *Hin*dIII library (0.06% of the library) were located in members of a LINE-like non-LTR retrotransposon family. Three and six members of this non-LTR retrotransposon were also present in the *Sal*I (0.05% of the library) and *Pst*I (0.06% of the library) libraries. It is well known that LTR retrotransposons are predominantly embedded in other repeats (SanMiguel et al. 1996; Wicker et al. 2001). The some tenfold higher prevalence of the $(CAA)_n$-containing LTR retroelement in the *Hin*dIII compared to the *Sal*I and *Pst*I libraries is thus in agreement with expectations. The similar frequency of the $(CAA)_n$-containing LINE-like element in the three libraries suggests that LINEs, in contrast to LTR retrotransposons, are preferentially associated with genes. This has also been observed in maize (C. Vitte and J. L. Bennetzen, unpublished data). When $(CAA)_n$ repeats in these two retrotransposon families are discounted, the overall frequency of SSRs in the *Sal*I and *Pst*I libraries is some twofold higher than in the *Hin*dIII library (Table 2). This confirms that SSRs, in particular dinucleotide repeats, in finger millet are preferentially present in genic, rather than intergenic DNA.

SSRs allocated to genes

A BLASTx search of the SSR-containing sequences against GenBank proteins located five dinucleotide repeats to introns, two di- and one trinucleotide repeats to 5′UTRs, four dinucleotide repeats to 3′UTRs, and three trinucleotide repeats to coding regions. In *Arabidopsis* and rice, $(GA)_n$ SSRs are significantly more frequent in 5′UTRs compared to introns and 3′UTRs (Morgante et al. 2002; Lawson and Zhang 2006). Such a trend was not seen in our data set. This may be due to the small number of SSRs that could unambiguously be allocated to genes. However, it should be noted that the stringency for declaring a SSR (≥10 repeat units for a dinucleotide repeat) is higher in our study compared to the Morgante et al. (2002) (≥6 dinucleotide repeats) and Lawson and Zhang (2006) studies (≥5 dinucleotide repeats). The average length of dinucleotide SSRs ≥ 12 bp was 16.1 bp, and 15.3 bp in *Arabidopsis* and rice ESTs, respectively (Morgante et al. 2002; La Rota et al. 2005). This is lower than the SSR threshold used in our study. As the relative proportion of SSR motifs varies with repeat length (La Rota et al. 2005), it is possible that their distribution is also length dependent. As expected, only trinucleotide repeats were found in coding regions.

Choice of a mapping population

When the finger millet mapping project was initiated, little was known about the genetics of the crop and polymorphism levels. In the early days, low levels of variation proved to be a serious constraint in the mapping of several plant species (e.g., in soybean, Apuya et al. 1988; wheat, Chao et al. 1989; and peanut, Kochert et al. 1991). A preliminary assessment of the levels of polymorphism within *E. coracana* subsp. *coracana* and between subsp. *coracana* and its wild progenitor subsp. *africana* was therefore carried out. Of twenty low-copy-number probes assessed with one restriction enzyme, none detected variation in a sample of four finger millet lines, three from different African countries (Ethiopia, Uganda, and Kenya) and one from Nepal (Table 1). Low levels of variation between finger millet accessions and an absence of a geographic bias had previously been reported by Salimath et al. (1995). The lack of variation between African and Asian finger millet lines might be attributable to the relatively recent transfer (an estimated 3,000 years ago) of finger millet from Africa, where the crop was domesticated some 5,000 years ago, to the Asian continent (de Wet et al. 1984). Furthermore, finger millet improvement has, until recently, been carried out mainly by pure line selection of local landraces. Pure lines may therefore represent only a fraction of the genetic diversity that is present in the finger millet landrace gene pool.

The diversity between the subspecies *coracana* and *africana* on the other hand is high. Seventy-five percent of probes assessed with one restriction enzyme showed variation between the two subspecies in one genome or the other. Hybrids between subsp. *coracana* and *africana* are fully fertile and are known to be formed where the two subspecies are sympatric (de Wet et al. 1984). We therefore opted to generate an inter-subspecific mapping population between the subsp. *africana* acc. MD-20 and subsp. *coracana* cv. Okhale-1. Depending on the marker system, between 62% (SSRs) and 70% (RFLPs and SSCP-ESTs) of the

markers tested could be mapped in the MD-20 × Okhale-1 population. The lower level of variation detected by the SSRs is likely due to the fact that the SSRs were mainly single copy and genome-specific, while the RFLP and SSCP-EST markers often hybridized/amplified across the A and B genomes of finger millet. Because most RFLP and SSCP-EST markers generally detected duplicated loci, their chances of detecting variation was twice as high as for the single copy SSR markers.

The finger millet genetic map

The genetic map incorporates homologous and heterologous RFLP, AFLP, EST, and SSR markers, and consists of 22 linkage groups, arranged in nine homoeology groups, plus four small linkage groups of unknown origin. Each of the nine homoeology groups contains an A and B genome linkage group, which likely correspond to the 18 finger millet chromosomes ($2n = 4x = 36$). The number of markers mapped is not significantly different between the A and B genomes. The A genome maps span 721 cM in 16 linkage groups and the B genome maps span 787 cM in nine linkage groups.

Marker distribution across the maps is fairly even, with some small clusters of markers that may correspond to regions of low recombination, for example around the centromeres (Chao et al. 1989; Devos et al. 1992; Qi et al. 1996, 2004). The relatively even distribution of markers is at least partly due to some level of marker selection carried out during the mapping process. The RFLP, AFLP, and SSR markers and the ESTs that were used during the initial mapping phase were selected for mapping only on the basis of their display of variation between MD-20 and Okhale-1, the parents of the mapping population. To obtain a finger millet map that was as complete as possible, the rice genome was used as a template for genome coverage. Once rudimentary maps had been obtained, additional ESTs for mapping were selected, based on their location in the rice genome (Srinivasachary and K. M. Devos, unpublished data).

The structure of homoeologous A and B genome chromosomes is highly similar. Our data provide no evidence that large chromosomal rearrangements have taken place during the recent evolution of either of the genomes. The only inconsistency in marker orders between A and B genome chromosomes was found between 3A and 3B (Fig. 1). The loci *Xlfo112* and *Xugep3* are closely linked on chromosome 3A but are 64 cM apart on 3B. In addition, *Xpsr104.2* is located in the *Xlfo112.2–Xugep3.1* interval on 3B, while the homoeologous locus is located on a linkage group that is independent from the 3A linkage group that carries *Xlfo112.1* and *Xugep3.2*. Provided that the loci *Xlfo112.1* and *Xlfo112.2* are truly orthologous, their different locations suggest the presence of a rearrangement in 3A compared to 3B. This rearrangement could span a large chromosome segment, or might have involved only a single gene. Single gene translocations have been frequently observed in comparative analysis of orthologous BAC clones of different grass species (Kilian et al. 1995; Foote et al. 1997; Tikhonov et al. 1999; Lai et al. 2004).

Conclusions

The genetic maps of finger millet, covering the nine homoeologous chromosome groups, provide an important first tool for genetic analysis of the crop and its improvement using marker-assisted breeding. Because the variation within cultivated finger millet is low, large numbers of highly variable markers such as SSRs will be required for breeding applications. Our data show that small-insert genomic libraries constructed using methylation-sensitive restriction enzymes are a good source of dinucleotide SSRs in finger millet. Building on this experience, we are in the process of developing a further set of SSRs to add to the 82 SSR markers described here. These resources are all available in the public domain and will aid in enhancing finger millet germplasm for critical traits such as blast resistance, lodging resistance, drought tolerance, and nutritional value.

Acknowledgments MMD and S gratefully acknowledge PhD studentships from the Sir Halley Stewart Fund (UK) and Commonwealth Scholarship Commission (UK), respectively. Part of the work was funded by a grant from the McKnight Foundation and a training grant from the Kirkhouse Trust (UK) to KMD.

References

Apuya NR, Frazier BL, Keim P, Jill Roth E, Lark KG (1988) Restriction fragment length polymorphisms as genetic markers in soybean, *Glycine max* (L.) Merrill. Theor Appl Genet 75:889–901

Barbeau WE, Hilu KW (1993) Protein, calcium, iron and amino acid content of selected wild and domesticated cultivars of finger millet. Plant Foods Hum Nutr 43:97–104

Bennetzen JL, Schrick K, Springer PS, Brown WE, SanMiguel P (1994) Active maize genes are unmodified and flanked by diverse classes of modified, highly repetitive DNA. Genome 37:565–576

Bisht MS, Mukai Y (2001) Genomic in situ hybridization identifies genome donor of finger millet (*Eleusine coracana*). Theor Appl Genet 102:825–832

Chao S, Sharp PJ, Worland AJ, Warham EJ, Koebner RMD, Gale MD (1989) RFLP-based genetic maps of wheat homoeologous group 7 chromosomes. Theor Appl Genet 78:495–504

de Wet JMJ, Prasada Rao KE, Brink DE, Mengesha MH (1984) Systematics and evolution of *Eleusine coracana* (*Gramineae*). Am J Bot 71:550–557

Devos KM, Atkinson MD, Chinoy CN, Liu C, Gale MD (1992) RFLP based genetic map of the homoeologous group 3 chromosomes of wheat and rye. Theor Appl Genet 83:931–939

Foote T, Roberts M, Kurata N, Sasaki T, Moore G (1997) Detailed comparative mapping of cereal chromosome regions corresponding to the *Ph1* locus in wheat. Genetics 147:801–807

Gruenbaum Y, Navey-Many T, Cedar H, Razin A (1981) Sequence specificity of methylation in higher plant DNA. Nature 292:860–862

Hilu KW, Johnson JL (1992) Ribosomal DNA variation in finger millet and wild species of *Eleusine* (Poaceae). Theor Appl Genet 83:895–902

Hiremath SC, Salimath SS (1991) Quantitative nuclear DNA changes in *Eleusine* (Gramineae). Plant Syst Evol 178:225–233

Hiremath SC, Salimath SS (1992) The A genome donor of *Eleusine coracana* Gaertn (Gramineae). Theor Appl Genet 84:747–754

Kilian A, Kudrna DA, Kleinhofs A, Yano M, Kurata N, Steffenson B, Sasaki T (1995) Rice-barley synteny and its application to saturation mapping of the barley *Rpg1* region. Nucleic Acids Res 23:2729–2733

Kochert G, Halward T, Branch WD, Simpson CE (1991) RFLP variability in peanut (*Arachis hypogaea* L.) cultivars and wild species. Theor Appl Genet 81:565–570

La Rota M, Kantety R, Yu Y-K, Sorrells ME (2005) Nonrandom distribution and frequencies of genomic and EST-derived microsatellite markers in rice, wheat, and barley. BMC Genomics 6:23–34

Lagercrantz U, Ellegren H, Andersson L (1993) The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. Nucleic Acids Res 21:1111–1115

Lai J, Ma J, Swigonova Z, Ramakrishna W, Linton E, Llaca V, Tanyolac B, Park Y-J, Jeong O-Y, Bennetzen JL, Messing J (2004) Gene loss and movement in the maize genome. Genome Res 14:1024–1931

Lawson M, Zhang L (2006) Distinct patterns of SSR distribution in the *Arabidopsis thaliana* and rice genomes. Genome Biol 7:R14.1–R14.11

Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nat Genet 30:194–200

Muza FR, Lee DJ, Andrews DJ, Gupta SC (1995) Mitochondrial DNA variation in finger millet (*Eleusine coracana*). Euphytica 81:199–205

Mysore KS, Baird V (1997) Nuclear DNA content in species of *Eleusine* (*Gramineae*): a critical re-evaluation using laser flow cytometry. Plant Syst Evol 207:1–11

National Research Council (1996) Lost crops of Africa; Volume I Grains. National Academy Press, Washington, DC

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

Qi X, Stam P, Lindhout P (1996) Comparison and integration of four barley genetic maps. Genome 39:379–394

Rachie KO, Peters LV (1977) The *Eleusines*. A review of World Literature. ICRISAT, Nairobi

Salimath SS, de Oliveira AC, Godwin ID, Bennetzen JL (1995) Assessment of genome origins and genetic diversity in the genus *Eleusine* with DNA markers. Genome 38:757–763

SanMiguel P, Tikhonov A, Jin Y-K, Motchoulskaia N, Zakharov D, Melake Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z, Bennetzen JL (1996) Nested retrotransposons in the intergenic regions of the maize genome. Science 274:765–768

Tikhonov AP, SanMiguel PJ, Nakajima Y, Gorenstein NM, Bennetzen JL, Avramova Z (1999) Colinearity and its exceptions in orthologous *adh* regions of maize and sorghum. Proc Natl Acad Sci 96:7409–7414

Vadivoo AS, Joseph R, Ganesan NM (1998) Genetic variability and diversity for protein and calcium contents in finger millet (*Eleusine coracana* (L.) Gaertn) in relation to grain colour. Plant Foods Hum Nutr 52:353–364

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

Werth CR, Hilu KW, Langner CA (1994) Isozymes of *Eleusine* (Gramineae) and the origin of finger millet. Am J Bot 81:1186–1197

Wicker T, Stein N, Albar L, Feuillet C, Schlagenhauf E, Keller B (2001) Analysis of a contiguous 211 kb sequence in diploid wheat (*Triticum monococcum* L.) reveals multiple mechanisms of genome evolution. Plant J 26:307–316