

HOSTED BY



ELSEVIER

Contents lists available at ScienceDirect

Electronic Journal of Biotechnology



Compilation of an informative microsatellite set for genetic characterisation of East African finger millet (*Eleusine coracana*)

Santie M. De Villiers^{a,c,*}, Vincent N. Michael^a, Eric O. Manyasa^a, Annis N. Saiyiorri^a, Santosh Deshpande^b

^a ICRISAT-Nairobi, PO Box 39063-00623, Nairobi, Kenya

^b ICRISAT-India, Patancheru 502 324, Andhra Pradesh, India

^c Pwani University, PO Box 195-80108, Kilifi, Kenya

ARTICLE INFO

Article history:

Received 20 June 2014

Accepted 6 November 2014

Available online xxx

Keywords:

Diversity analysis

SSR markers

Genotyping

ABSTRACT

Background: Genetic diversity of finger millet (*Eleusine coracana*), a nutritious neglected staple cereal in Africa and South Asia is largely uncharacterized. This study analysed 82 published SSR markers for finger millet across 10 diverse accessions to compile an informative set for genetic characterisation. Extensive optimization compared single samples with bulked leaf or bulked DNA samples for capturing within accession genetic diversity. The markers were evaluated to determine (1) how efficiently they amplified target loci during high-throughput genotyping with a generic PCR protocol, (2) ease of scoring PCR products and (3) polymorphism and ability to discern genetic diversity within the tested finger millet germplasm.

Results: Across 88 samples, the 52 markers that worked well amplified 274 alleles, ranging from 2 to 14 per locus with a mean of 4.89. Major allele frequency ranged from 0.18 to 0.93 with a mean of 0.57. Polymorphic Information Content (PIC) ranged from 0.13 to 0.88 with a mean of 0.5 and availability varied between 64 and 100% with a mean of 92.8%. Heterozygosity ranged from 0 to 1.0, with a mean of 0.26.

Discussion: Five individual samples from an accession captured the largest number of alleles per locus compared to the four different bulked sampling strategies but this difference was not significant. The identified set comprised 20 markers: UGEP24, UGEP53, UGEP84, UGEP27, UGEP98, UGEP95, UGEP64, UGEP33, UGEP67, UGEP106, UGEP110, UGEP57, UGEP96, UGEP66, UGEP46, UGEP79, UGEP20, UGEP12, UGEP73 and UGEP5 and was since used to assess East African finger millet genetic diversity in two separate studies.

© 2014 Pontificia Universidad Católica de Valparaíso. Production and hosting by Elsevier B.V. All rights reserved.

1. Introduction

Finger millet (*Eleusine coracana* $2n = 4x = 36$) is a nutritious cereal staple that originated in the tropical and sub-tropical parts of Africa and Asia. It is adapted to a wide range of adverse agro-ecological conditions with minimal inputs and provides critical plant genetic resources upon which millions depend for food and rural household incomes on infertile and marginal lands [1]. Finger millet ranks third in cereal production in the semi-arid regions of the world after sorghum and pearl millet [2] but remains one of the neglected and underutilized crops of Africa even though it is extensively cultivated, mostly by subsistence farmers. It serves as a food security crop because of its high nutritional value and excellent storage qualities [3]. The seeds contain up to 14% protein and are particularly rich in methionine, iron, and calcium and in marginal and medium agricultural zones in Ethiopia, Eritrea and Uganda; this crop is a high priority staple [4].

To date, a limited number of SSR markers have been reported for finger millet. In this study, we analysed the 82 markers developed by Dida et al. [3] for their usefulness in discerning the genetic diversity in finger millet, using a selection of 10 diverse cultivated finger millet accessions from the mini-core collection developed by Upadhyaya et al. [5].

An extensive optimization study was undertaken to determine if DNA from more than one individual can be bulked in a single sample in order to reduce the number of samples needed for genotyping a single accession. Bulking would allow for the cost-effective analysis of a larger number of genotypes, rather than analysing multiple individuals from a single accession. This is a feasible approach for finger millet as the inherent nature of the crop as an allotetraploid, is 99% inbred [6], and in general it behaves like a diploid crop species, with 2 homozygous alleles per locus and little variation within an accession on the DNA level expected. Alternatively we wanted to determine if a single sample from an accession is adequate to capture within accession diversity. To this end, the 82 SSR markers were evaluated on how well they worked in a high-throughput genotyping system, specifically in terms of the use of a generic PCR protocol, ease of scoring PCR products, polymorphism and ability to discern genetic diversity within the tested finger millet germplasm.

* Corresponding author.

E-mail address: santiedevilliers@gmail.com (S.M. De Villiers).

Peer review under responsibility of Pontificia Universidad Católica de Valparaíso.

<http://dx.doi.org/10.1016/j.ejbt.2014.12.001>

0717-3458/© 2014 Pontificia Universidad Católica de Valparaíso. Production and hosting by Elsevier B.V. All rights reserved.

Please cite this article as: De Villiers SM, et al, Compilation of an informative microsatellite set for genetic characterisation of East African finger millet (*Eleusine coracana*), (2015), <http://dx.doi.org/10.1016/j.ejbt.2014.12.001>

2. Materials and methods

2.1. DNA extraction

Ten diverse finger millet accessions were selected to represent the diverse agro-ecologies where finger millet is grown in its main area of cultivation in eastern Africa in Kenya, Uganda and Tanzania, listed in Table 1. They comprised of accessions GBK-000414A, GBK-011135A and GBK-044047A from Kenya obtained through the National Genebank of Kenya (NGBK), Sansamula and Namakonta from Tanzania, Ebega, Emorumu and Bulu from Uganda and IE2572 and IE2957 from the international mini-core collection for finger millet [5]. For each accession, DNA was extracted from fresh leaves of 14-d-old seedlings of 5 individuals to prepare 9 different samples for genotyping. These comprised of DNA from each of the 5 individuals separately (samples 1 to 5) as well as equal sized leaf samples from 3 individuals bulked (3 LB – sample 6) or leaves from 5 individuals bulked (5 LB – sample 7) in a single sample. In addition, extracted DNA, normalized at 10 ng/μL post-extraction, was pooled from 3 individuals (3 DB – sample 8) and 5 individuals (5 DB – sample 9). DNA extraction was done according to the modified CTAB protocol of Mace et al. [7], omitting the phenol:chloroform step. Extracted DNA was visualised on a 0.8% (w/v) agarose gel and quantified spectrophotometrically using a Nanodrop® 1000 (Thermo Scientific, USA), followed by dilution to 10 ng/μL in TE buffer (10 mM Tris, 0.1 mM EDTA pH 8.0).

2.2. SSR genotyping optimization

The 9 different DNA samples described above for each accession (90 samples in all) were subjected to genotyping using 82 published SSR markers for finger millet [3]. All forward primers contained an M13-tag (5'-CACCACGTTGTAACGAC-3') on the 5' end that allowed incorporation of a fluorescent label during PCR to allow detection of amplification products [8]. PCR amplification was performed in 10 μL in 384 well microtitre plates and each reaction comprised of 1 × PCR buffer (20 mM Tris-HCl, pH 7.6; 100 mM KCl; 0.1 mM EDTA; 1 mM DTT; 0.5% (w/v) Triton X-100; 50% (v/v) glycerol), 2 mM MgCl₂, 0.16 mM dNTPs, 0.16 μM fluorescent labelled M13-forward primer, 0.04 μM forward primer, 0.2 μM reverse primer, 0.2 units of Taq DNA polymerase (SibEnzyme Ltd, Russia) and 30 ng of template DNA. PCR reactions were performed on a GeneAmp 9700 thermocycler (Applied Biosystems) with initial denaturation of 94°C for 5 min, followed by 35 cycles of 94°C for 30 s, 59°C for 1 min and 72°C for 2 min, followed by final elongation at 72°C for 20 min. Amplification was confirmed by running 4 μL of the products on a 2% (w/v) agarose gel stained with GelRed® (Biotium, USA) and visualised under UV light.

Amplification products (1.5 μL–3.5 μL of each) were co-loaded in sets of 3 to 4 markers together with the internal size standard, GeneScan™–500 LIZ® (Applied Biosystems) and Hi-Di™ Formamide

(Applied Biosystems) and separated by capillary electrophoresis using an ABI Prism® 3730 Genetic analyzer (Applied Biosystems). Allele calling was performed with Gene Mapper 4.0 (Applied Biosystems) and allelic data for each marker analysed with PowerMarker V3.25 [9] and Arlequin [10]. DARwin V5 [11] was used to generate unweighted neighbour-joining dendrograms from the allelic data of the 52 SSRs that worked well as well as a subset of 20 markers that was selected for use in further genetic diversity studies on finger millet. This was done to confirm that the selected subset could discern genetic diversity as effectively as the larger subset of 53 markers.

3. Results

3.1. DNA extraction and PCR genotyping

Good quality and amounts of DNA, confirmed by 0.8% (w/v) agarose gel electrophoresis and spectrophotometry were obtained for all the samples and all except 408_4 and 5_4 worked well in subsequent PCR. These two samples were removed from the dataset so that the final number analysed was 88. Generally, PCR products were 19 bp longer than the expected amplified fragment size due to the incorporation of the 5'-M13 tag [8].

In the current study 13 of the 82 markers were monomorphic (UGEP7, UGEP11, UGEP26, UGEP34, UGEP50, UGEP54, UGEP62, UGEP69, UGEP75, UGEP88, UGEP91, UGEP93 and UGEP104). An additional 16 markers did not work well (UGEP1, UGEP6, UGEP8, UGEP9, UGEP10, UGEP13, UGEP16, UGEP22, UGEP25, UGEP29, UGEP65, UGEP86, UGEP90, UGEP101, UGEP107, UGEP108 and UGEP111) and UGEP108 was highly heterozygous as well as heterogeneous within accessions and therefore not informative. These markers were excluded from the data set and only markers that amplified identical and clearly discernable alleles in at least eight of the nine samples of each accession were further considered. Four markers (UGEP5, UGEP51, UGEP95 and UGEP103) appeared to amplify two loci each and were scored as two separate markers each. This resulted in a subset of 52 markers (listed in Table 2), amplifying 56 loci.

Across the 88 samples, the total number of alleles amplified by the 52 markers was 274, ranging from 2 to 14 alleles per locus with a mean of 4.89. The major allele frequency ranged from 0.18 to 0.93 with a mean of 0.57. The Polymorphic Information Contents (PIC) ranged from 0.13 to 0.88 with a mean of 0.5 across all loci and availability varied between 64 and 100% with a mean of 92.8%. Heterozygosity ranged from 0 to 0.94, with a mean of 0.25. Details of these results are presented in Table 2.

3.2. Bulk sample strategy

Bulking of samples was compared to results from individual samples using Arlequin software [10] and the results are presented in Table 3. Individual samples, where data from the 5 individuals per accession were combined, detected the largest number of alleles per locus 95% of the time. All of the pooling strategies detected the largest number of alleles less frequently than the combined individual samples. Bulk DNA samples were more effective than the leaf bulk strategy. Bulk DNA from 3 individuals (3 DB) detected the largest number of alleles 53% of the time but generally detected 86% of the all the alleles that individual samples did. DNA from 5 samples bulked (5 DB) and bulked leaf samples from 3 individuals (3 LB) detected the largest number of alleles 50% of the time (87 and 86% of all alleles detected by individual samples, respectively) and 5 leaf samples bulked (5 LB) did so 43% of the time, detecting 83% of the alleles that individual samples did.

The various sampling strategies were also compared for the effectiveness of detecting the major allele within each accession using PowerMarker [9] and all strategies detected the major allele between

Table 1
Passport data of the accessions used in this study.

Sample No.	Acc/local name	Country	District	Comment
253	GBK-044047A	Kenya	Laikipia	High altitude
124	GBK-000414A	Kenya	South Nyanza	Medium altitude
169	GBK-011135A	Kenya	Machakos	Lowland
386	Sansamula	Tanzania	Sumbawanga	Medium altitude
408	Namakonta	Tanzania	Sumbawanga	High altitude
5	Ebega	Uganda	Sererere	Eastern Uganda
77	Bulu	Uganda	Hoima	Eastern Uganda
33	Emorumu (rock)	Uganda	Amuria	Western Uganda
271	IE2572	Kenya	Mini-core collection	Late maturing
281	IE2957	Germany	Mini-core collection	Late maturing

Table 2

Details of the 52 markers, arranged according to PIC that worked well in this study and was assessed to compile a genotyping tool kit. The first 20 markers comprise the tool kit.

Marker	Repeat motif	Allele size	No of alleles	MAF (%)	PIC	Gene diversity	Avail (%)	H	LG	
1	UGEP24	(GA) ₂₆	171–223	14	20	0.88	0.89	98	0.31	3B
2	UGEP53	(AG) ₂₆	211–257	9	18	0.87	0.88	98	0.25	2A
3	UGEP84	(CT) ₂₄	166–192	8	23	0.82	0.84	98	0.23	–
4	UGEP27	(GA) ₁₉	267–285	8	23	0.81	0.84	98	0.09	–
5	UGEP98	(GCC) ₈	181–282	9	26	0.81	0.83	93	0.14	–
6	UGEP95_2	(TC) ₁₄	235–245	8	23	0.79	0.81	98	0.36	–
7	UGEP64	(CT) ₂₃	233–245	11	37	0.78	0.80	94	0.29	–
8	UGEP33	(TC) ₁₈	211–237	7	35	0.76	0.79	93	0.81	–
9	UGEP67	(TC) ₂₂ TT(GT) ₅	230–252	8	34	0.76	0.78	96	0.03	–
10	UGEP106	(AC) ₁₂	182–190	7	28	0.75	0.79	89	0.25	9B
11	UGEP110	(CT) ₁₂	204–216	6	31	0.74	0.78	86	0.88	–
12	UGEP57	(AG) ₁₆	467–473	6	29	0.73	0.77	92	0.16	–
13	UGEP96	(CT) ₁₀	246–310	8	38	0.72	0.76	64	0.45	–
14	UGEP66	(AG) ₂₉	211–237	9	46	0.72	0.74	93	0.10	–
15	UGEP46	(GA) ₁₄	179–191	6	37	0.70	0.75	83	0.07	–
16	UGEP79	(CT) ₁₂	240–258	8	43	0.70	0.74	97	0.06	–
17	UGEP20	(GA) ₂₀	146–158	5	33	0.69	0.74	100	0.84	–
18	UGEP12	(CT) ₂₂	212–250	7	37	0.69	0.74	97	0.20	8B
19	UGEP73	(CT) ₄ CC(CT) ₁₀	176–194	6	42	0.67	0.71	92	0.27	7AB
20	UGEP5_1	(TC) ₁₂ AC(TC) ₄	224–244	4	37	0.66	0.71	91	0.66	9B
21	UGEP74	(CT) ₂₁	214–230	5	44	0.64	0.69	96	0.14	–
22	UGEP59	(GA) ₁₁	254–264	5	48	0.61	0.67	97	0.85	–
23	UGEP15	(CT) ₂₂	195–203	6	56	0.60	0.64	98	0.13	3Aa
24	UGEP47	(GA) ₁₈	216–222	4	39	0.60	0.66	68	0.59	–
25	UGEP56	(GT) ₁₂	175–183	4	52	0.54	0.61	98	0.94	9A
26	UGEP70	(TCC) ₇	249–255	3	49	0.51	0.60	98	0.05	–
27	UGEP28	(TC) ₁₇	238–250	5	63	0.51	0.54	100	0.03	–
28	UGEP21	(GA) ₁₆	210–220	3	50	0.46	0.50	100	0.00	1B
29	UGEP31	(GA) ₁₂	254–260	4	67	0.45	0.55	97	0.10	3Aa
30	UGEP105	(GA) ₂₀	188–200	4	57	0.42	0.52	90	0.80	–
31	UGEP3	(TC) ₁₂ AC(TC) ₄	220–222	3	57	0.40	0.51	94	0.06	3ABb
32	UGEP68	(CT) ₁₄	254–258	3	73	0.38	0.43	93	0.05	9B
33	UGEP60	(GA) ₃₇	194–214	2	50	0.38	0.50	98	0.11	1B
34	UGEP103_1	(GTT) ₈	184–198	2	57	0.37	0.49	94	0.86	–
35	UGEP78	(GA) ₁₄	242–252	6	77	0.37	0.39	94	0.15	2B
36	UGEP80	(CAT) ₇	202–205	2	63	0.36	0.47	97	0.11	–
37	UGEP81	(CT) ₁₁ /(CT) ₈	287–295	3	71	0.35	0.42	97	0.24	6B
38	UGEP95_1	(TC) ₁₄	225–229	2	66	0.35	0.45	96	0.65	–
39	UGEP58	(CA) ₁₅	208–214	3	74	0.34	0.40	91	0.01	–
40	UGEP97	(GT) ₈ TT(GT) ₁₃	173–175	2	69	0.33	0.42	80	0.00	–
41	UGEP51_2	(GT) ₁₁	316–336	3	79	0.33	0.35	96	0.00	–
42	UGEP51_1	(GT) ₁₁	210–230	4	80	0.32	0.35	98	0.00	–
43	UGEP5_2	(TC) ₁₂ AC(TC) ₄	316–336	3	79	0.32	0.35	76	0.12	9B
44	UGEP18	(CT) ₁₂	239–249	3	80	0.31	0.33	88	0.09	1B
45	UGEP45	(GA) ₁₃	233–237	5	80	0.31	0.31	66	0.12	–
46	UGEP52	(GA) ₁₆	227–231	3	82	0.29	0.32	98	0.00	2A
47	UGEP19	(GA) ₁₈	239–249	3	83	0.27	0.29	94	0.08	–
48	UGEP109	(GA) ₁₃	170–250	3	82	0.26	0.30	76	0.09	–
49	UGEP83	(GA) ₂₆	209–211	2	82	0.25	0.30	96	0.08	–
50	UGEP100	(CT) ₁₂	253–262	3	85	0.24	0.26	96	0.03	–
51	UGEP76	(CT) ₁₉	205–215	2	85	0.22	0.25	97	0.30	3B
52	UGEP77	(GA) ₁₄	216–222	3	87	0.21	0.23	96	0.13	4B
53	UGEP102	(TG) ₁₇	170–198	4	89	0.19	0.20	80	0.19	10
54	UGEP 17	(GTT) ₇	237–240	2	90	0.17	0.18	98	0.00	–
55	UGEP103_2	(GTT) ₈	206–214	3	92	0.14	0.15	83	0.08	–
56	UGEP87	(TC) ₃₉	152–156	3	93	0.13	0.14	97	0.13	–
	Mean		4.89	57	0.50	0.54	92	0.25		
	Min		2	18	0.13	0.14	64	0.00		
	Max		14	93	0.88	0.89	100	0.94		
	Total		274							

MAF = Major Allele Frequency; PIC = Polymorphic Information Contents; Avail = Availability; H = Heterozygosity; LG = Linkage group where marker has been mapped (Dida et al. [3]).

93 and 100% of the time. 3LB was most effective at 100%, followed by 5LB (98%), 5DB (96%), 3DB (93%) and individual samples (93%).

3.3. Phylogenetic analysis

Fig. 1 depicts the unweighted neighbour-joining dendrograms of the allelic data from the 52 SSRs (Fig. 1a) that worked well as well as the subset of 20 markers (Fig. 1b) that was selected for use in further genetic diversity studies on finger millet. The samples from each of the ten accessions were grouped into tight individual clusters in

both dendrograms, with the exception of Acc 408_2 in Fig. 1a that was loosely associated with Acc 386.

4. Discussion

The PCR protocol reported here is used routinely in the laboratory at ICRISAT-Nairobi and generally works well for high throughput SSR genotyping with sorghum, pigeonpea and groundnuts, even though it did not work well for 16 of the 82 markers tested in this study. However, since it is important to have a standardised protocol that works well across all DNA samples, it was decided to focus only on the

Table 3
Summary of sample by marker, indicating which bulked samples captured the largest numbers of alleles that were contributed by the sum of all the individuals within an accession.

Locus	Number of alleles detected by each sampling strategy								% of total no of alleles detected by each sampling strategy					% of total no of alleles detected compared to individual samples				
	3 DB	3 LB	5 DB	5 LB	Ind	Mean	s.d.	Total no	3 DB	3 LB	5 DB	5 LB	Ind	3 DB	3 LB	5 DB	5 LB	Ind
UGEP3	3	2	3	3	3	2.8	0.40	4	75	50	75	75	75	100	67	100	100	100
UGEP5_1	4	4	4	3	4	3.8	0.40	5	80	80	80	60	80	100	100	100	75	100
UGEP5_2	3	3	3	2	3	2.8	0.40	4	75	75	75	50	75	100	100	100	67	100
UGEP12	4	4	5	5	7	5.0	1.10	8	50	50	63	63	88	57	57	71	71	100
UGEP15	4	6	5	5	6	5.2	0.75	7	57	86	71	71	86	67	100	83	83	100
UGEP17	2	2	2	2	2	2.0	0.00	3	67	67	67	67	67	100	100	100	100	100
UGEP18	3	2	3	3	3	2.8	0.40	4	75	50	75	75	75	100	67	100	100	100
UGEP19	3	2	2	2	3	2.4	0.49	4	75	50	50	50	75	100	67	67	67	100
UGEP20	5	4	5	5	5	4.8	0.40	5	100	80	100	100	100	100	80	100	100	100
UGEP21	3	3	3	3	3	3.0	0.00	3	100	100	100	100	100	100	100	100	100	100
UGEP24	11	11	11	11	14	11.6	1.20	15	73	73	73	73	93	79	79	79	79	100
UGEP27	8	7	5	6	8	6.8	1.17	9	89	78	56	67	89	100	88	63	75	100
UGEP28	3	4	4	4	5	4.0	0.63	5	60	80	80	80	100	60	80	80	80	100
UGEP31	4	4	4	3	4	3.8	0.40	5	80	80	80	60	80	100	100	100	75	100
UGEP33	7	7	6	7	7	6.8	0.40	8	88	88	75	88	88	100	100	86	100	100
UGEP45	2	3	2	3	4	2.8	0.75	6	33	50	33	50	67	50	75	50	75	100
UGEP46	5	4	4	6	6	5.0	0.89	7	71	57	57	86	86	83	67	67	100	100
UGEP47	4	4	4	3	4	3.8	0.40	5	80	80	80	60	80	100	100	100	75	100
UGEP51_1	3	3	3	3	4	3.2	0.40	5	60	60	60	60	80	75	75	75	75	100
UGEP51_2	3	3	3	3	3	3.0	0.00	4	75	75	75	75	75	100	100	100	100	100
UGEP52	3	3	3	3	3	3.0	0.00	4	75	75	75	75	75	100	100	100	100	100
UGEP53	9	9	8	9	9	8.8	0.40	10	90	90	80	90	90	100	100	89	100	100
UGEP56	3	3	4	3	3	3.2	0.40	5	60	60	80	60	60	100	100	133	100	100
UGEP57	6	6	5	5	6	5.6	0.49	7	86	86	71	71	86	100	100	83	83	100
UGEP58	2	2	2	2	3	2.2	0.40	4	50	50	50	50	75	67	67	67	67	100
UGEP59	4	4	4	4	5	4.2	0.40	6	67	67	67	67	83	80	80	80	80	100
UGEP60	2	2	2	2	2	2.0	0.00	3	67	67	67	67	67	100	100	100	100	100
UGEP64	8	8	8	8	11	8.6	1.20	12	67	67	67	67	92	73	73	73	73	100
UGEP66	8	8	8	8	9	8.2	0.40	10	80	80	80	80	90	89	89	89	89	100
UGEP67	7	6	6	6	8	6.6	0.80	9	78	67	67	67	89	88	75	75	75	100
UGEP68	3	3	3	3	3	3.0	0.00	4	75	75	75	75	75	100	100	100	100	100
UGEP70	3	3	3	3	3	3.0	0.00	4	75	75	75	75	75	100	100	100	100	100
UGEP73	4	5	5	5	6	5.0	0.63	7	57	71	71	71	86	67	83	83	83	100
UGEP74	4	4	4	4	5	4.2	0.40	6	67	67	67	67	83	80	80	80	80	100
UGEP76	2	2	2	2	2	2.0	0.00	3	67	67	67	67	67	100	100	100	100	100
UGEP77	2	2	1	2	3	2.0	0.63	4	50	50	25	50	75	67	67	33	67	100
UGEP78	3	3	4	4	6	4.0	1.10	7	43	43	57	57	86	50	50	67	67	100
UGEP79	4	4	5	2	7	4.4	1.63	9	44	44	56	22	78	57	57	71	29	100
UGEP80	2	2	2	2	2	2.0	0.00	3	67	67	67	67	67	100	100	100	100	100
UGEP81	2	3	3	2	3	2.6	0.49	4	50	75	75	50	75	67	100	100	67	100
UGEP83	2	2	2	2	2	2.0	0.00	3	67	67	67	67	67	100	100	100	100	100
UGEP84	8	8	8	8	8	8.0	0.00	9	89	89	89	89	89	100	100	100	100	100
UGEP87	2	2	2	1	3	2.0	0.63	4	50	50	50	25	75	67	67	67	33	100
UGEP95_1	2	2	2	2	2	2.0	0.00	3	67	67	67	67	67	100	100	100	100	100
UGEP95_2	6	5	5	5	8	5.8	1.17	9	67	56	56	56	89	75	63	63	63	100
UGEP96	4	6	2	6	7	5.0	1.79	9	44	67	22	67	78	57	86	29	86	100
UGEP97	2	2	2	2	2	2.0	0.00	3	67	67	67	67	67	100	100	100	100	100
UGEP98	8	8	6	8	9	7.8	0.98	10	80	80	60	80	90	89	89	67	89	100
UGEP100	3	3	3	2	3	2.8	0.40	4	75	75	75	50	75	100	100	100	67	100
UGEP102	1	2	2	1	3	1.8	0.75	4	25	50	50	25	75	33	67	67	33	100
UGEP103_1	2	2	2	2	2	2.0	0.00	3	67	67	67	67	67	100	100	100	100	100
UGEP103_2	2	1	3	1	3	2.0	0.89	4	50	25	75	25	75	67	33	100	33	100
UGEP105	2	3	3	2	3	2.6	0.49	5	40	60	60	40	60	67	100	100	67	100
UGEP106	4	5	5	6	6	5.2	0.75	8	50	63	63	75	75	67	83	83	100	100
UGEP109	2	2	3	2	2	2.2	0.40	4	50	50	75	50	50	100	100	150	100	100
UGEP110	6	5	5	5	5	5.2	0.40	7	86	71	71	71	71	120	100	100	100	100
Efficiency*									50	46	50	43	95					
Mean	3.95	3.96	3.89	3.9	4.7	4.1	0.33	5.82	67	67	67	65	79	86	86	87	83	100
s.d.	2.21	2.16	1.95	2.2	2.6	2.2	0.21	2.63										

Values in bold indicates the sample that captured the largest number of alleles within an accession.

* % times that the sample captures the highest no ($\leq 100\%$) of alleles.

markers that worked well with this protocol. Further elimination of the 13 monomorphic markers and one highly heterozygous and heterogeneous marker, resulted in a total of 52 markers that were assessed, of which four (UGEP5, UGEP51, UGEP95 and UGEP103) were scored as amplifying duplicate loci. This subset of 52 markers (listed in Table 2), amplifying 56 loci, was used to identify the best SSR kit for diversity assessment.

This study set out to identify a set of 20 SSR markers since reports on genetic diversity studies used sets of 15 to 50 SSR markers for germplasm collections ranging from a few hundred to thousands of accessions such as for sorghum (41 SSRs, 3367 accessions) [12] as well as for sets developed for global germplasm diversity assessment by the Generation Challenge Programme (GCP) for barley (15 SSRs, 2692 accessions), chickpea (35 SSRs, 3000 accessions), coconut (30

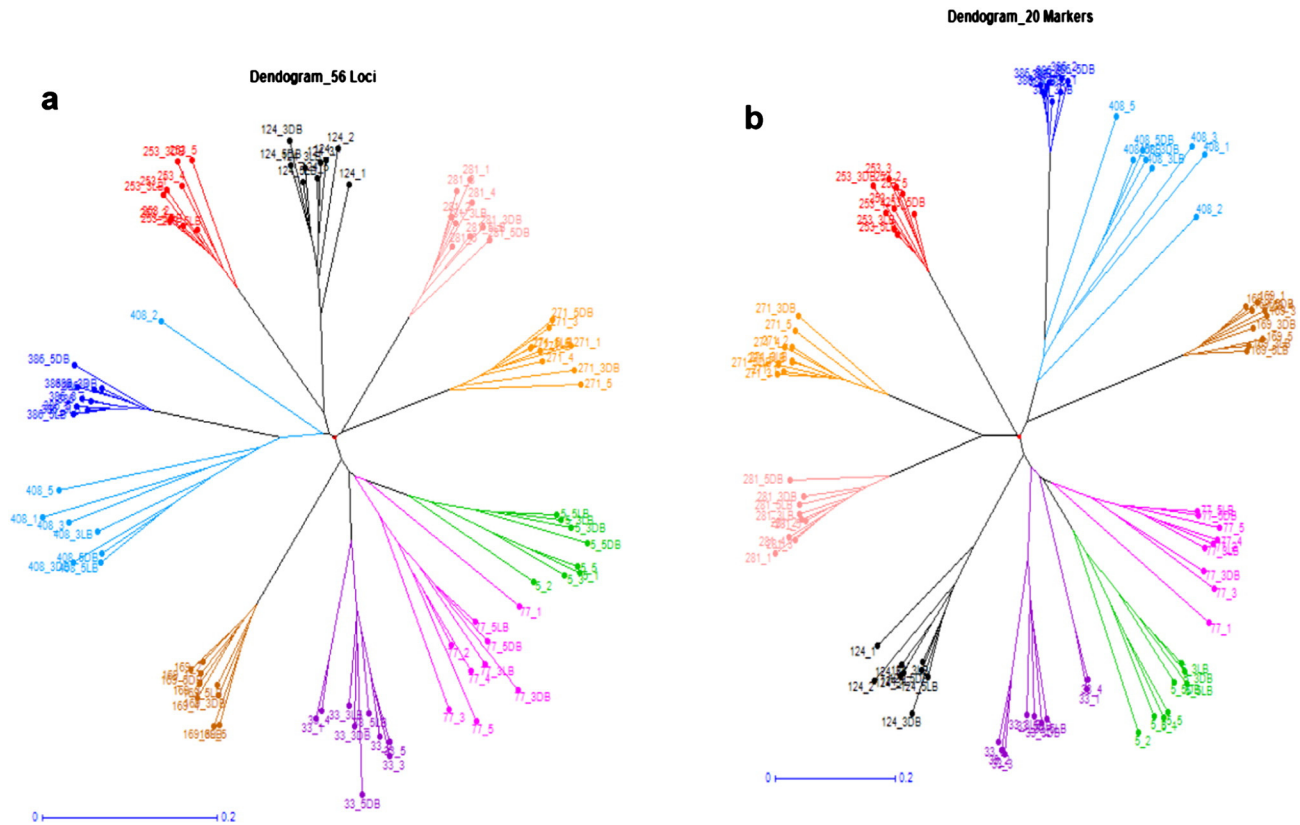


Fig. 1. Unweighted neighbour-joining dendrograms of the allelic data from (a) the 52 SSRs that worked well and (b) the subset of 20 markers that was selected for use in further genetic diversity studies on East African finger millet. Colour key: Acc 5 (green), Acc 33 (purple), Acc 77 (bright pink), Acc 124 (black), Acc 169 (brown), Acc 253 (red), Acc 271 (orange), Acc 281 (pale pink), Acc 386 (dark blue) and Acc 408 (light blue).

SSRs, 201 accessions), groundnuts (21 SSRs, 911 accessions), pearl millet (21 SSRs, 1000 accessions), pigeonpea (20 SSRs, 952 accessions) and pearl millet (20 SSRs, 1000 accessions) [13,14]. The sets with larger numbers of markers were developed for crops with relatively abundant genomic resources such as sorghum and chickpea compared to the more neglected and orphan crops such as pearl millet and pigeonpea, for which smaller sets of about 20 markers were used, likely because fewer markers were available for these crops to select from, as was also the case in this study with finger millet.

Table 2 indicates that across all samples, the total number of alleles amplified was 274 (2 to 14 alleles per marker, mean of 4.89). This was in line with the value reported by Arya et al. [15] of a mean of 4 alleles per locus for 17 SSRs but lower than the 6.42 for 79 SSRs for cultivated and wild *Eleusine* species from Africa and India reported by Dida et al. [16]. The major allele frequency, indicating how widely the dominant allele amplified by a particular marker is spread across the samples, ranged from 0.18 to 0.93 (0.57 mean) and PIC ranged from 0.13 to 0.88 (0.5 mean) across all markers. Gene diversity values obtained in this study (Mean of 0.55, ranging from 0.14 to 0.89) was also comparable with that obtained by Arya et al. [15] (mean of 0.47, ranging from 0 to 0.73) and Dida et al. [16] (mean of 0.47). Availability, an indication of how well a marker worked, ranged from 64 to 100% with a mean of 92.8%. These criteria were used to identify the 20 most suitable markers for genetic diversity assessment for finger millet.

4.1. Bulk sample strategy

Bulking of samples allow more cost-effective analysis of a larger number of genotypes, compared to multiple individuals from a single accession. This was considered a feasible approach for finger millet, which is 99% inbred and behaves like a diploid species [6], with 2

homozygous alleles per locus. Likewise, it was also considered whether a single sample from an accession could detect the predominant within accession diversity. Genotyping data for all samples were scored as diploid, selecting the 2 most prominent alleles in each sample. This strategy determined that, for bulked samples, private and rare alleles were largely eliminated and it also meant that these samples could not detect all the alleles in every sample.

Results showed that, as expected, individual samples (data from 5 individuals per accession combined) generally captured the largest number of alleles per locus 95% of the time and therefore displayed the maximum genetic diversity per accession. When considering the results from the bulked samples presented in Table 3, bulked DNA samples were marginally more effective than bulked leaves (86–87% compared to 83–86%) to capture the same number of alleles per locus as the combined 5 individual samples. All samples often exhibited fewer than the total no of alleles per accession due to the exclusion of rare and unique alleles from the dataset during both allele scoring and data curation, as explained above.

When the 5 individual samples from an accession was compared, on average each presented the major allele 93% of the time and bulked samples did so 95% (3 DB), 96% (5 DB), 98% (5 LB) and 100% (3 LB) of the time. Since we considered only the 2 most prominent alleles in each sample, the alleles that were excluded when bulked samples were compared to the individual samples, were the rare alleles that would not have contributed to the genetic diversity estimates within the population. 3 DB were found to be the best strategy to detect the maximum number of alleles and 3 LB to capture the major allele most often compared to individual samples, but these differences are not statistically meaningful. Therefore, depending on the available resources, either individual plants or bulked DNA samples can be used.

In bulked samples, it was noted that a minority of individuals (1 or 2) from an accession could present an allele different to that of the

majority in a group of 5. This complicated the scoring of the data since multiple alleles, often in odd numbers confound the data for a diploid or homozygous tetraploid crop. For the most part in this study, such a unique allele in an accession presented as a smaller peak in the GeneMapper profile as it was not equally contributed by all individuals and did not complicate the selection of the 2 major alleles. However, in some cases the PCR step may have amplified one locus preferentially compared to the others but since this could not be inferred from the GeneMapper profiles, in such cases, the two major alleles (highest peaks) were selected. For some markers (UGEP5, UGEP51, UGEP95 and UGEP103) there were clearly more than 2 alleles in the majority of samples, even for single samples. In these cases it was most likely due to either a heterozygous or duplicate locus that was amplified in this tetraploid crop. For these markers all genotypes were assessed to confirm that the multiple alleles occurred consistently. When this was the case, as for UGEP5, UGEP51, UGEP95 and UGEP103, the marker data were split and evaluated as two individual loci. However, when very few individuals presented such multiple alleles, a decision had to be made on which 2 alleles were the major alleles to be included in the data set. To ensure a robust set of markers, wherever possible, such ambiguous markers were avoided.

4.2. SSR tool kit identification

The allelic data from the selected 56 loci were analysed using PowerMarker V3.25 [9] and were ranked according to (1) their ability to discern amongst the different genotypes using the polymorphic information content (PIC), (2) availability, which indicated how often the marker worked across the 88 samples analysed, (3) number of alleles presented per marker and (4) the frequency of the major allele as main criteria. Highly heterozygous and heterogeneous markers were avoided. The best 20 SSR markers for finger millet genotyping are indicated in Table 2. A phylogenetic analysis was carried out with the allelic data for all 52 markers as well as for the 20 markers selected as the “genotyping kit”, presented in Fig. 1. As expected, the topography of the two dendrograms looked similar, indicating that the selected 20 markers represent the data from the entire set of 52 markers. Therefore, genetic analysis using the sub-set of 20 markers should capture the same genetic diversity as the 52 markers. It was interesting to note that this list differed substantially from the set of markers prescribed by the GCP Bioinformatics Central Registry [13] for finger millet genotyping (http://gcpcr.grinfo.net/index.php?app=datasets&inc=files_list). For the germplasm analysed in this study, several of the GCP markers were found to be either monomorphic, exhibited low PIC or did not work well.

5. Conclusions

The results from this study present the best set of microsatellite markers that could be assembled from the currently available finger millet SSR primers in the public domain that can be used for genetic diversity assessment of finger millet germplasm. One obvious shortcoming was that very few of these markers have been mapped [3] and therefore it was not clear how well they were spread across the genome as compared to the SSR reference set recently reported by Billot et al. [12]. Also, there were very few markers to work with from the outset, emphasizing the need for further investment in finger millet genomics to allow additional marker- and genetic resource development for this important orphan crop. This set of markers was subsequently used successfully to assess the genetic diversity amongst

340 finger millet accessions from Kenya, Tanzania and Uganda as reported by Manyasa et al. [17] as well as for 76 accessions comprising of cultivated (*E. coracana* subsp. *coracana*) and its wild relatives *Eleusine intermedia*, *Eleusine indica*, *Eleusine multiflora* and *Eleusine floccifolia* from Ethiopia [18].

Financial support

This study formed part of the project “Delivering New Sorghum and Finger Millet Innovations for Food Security and Improving Livelihoods in Eastern Africa” supported by the Swedish International Development Agency (SIDA) (Project no: 01/2010) Bio-Innovate programme.

References

- [1] Hilu KW, De Wet MJM, Harlan JR. Archaeobotanical studies of *Eleusine coracana* ssp. *coracana* (finger millet). *Am J Bot* 1979;66:330–3. <http://dx.doi.org/10.2307/2442610>.
- [2] Barbeau WE, Hilu KW. Protein, calcium, iron and amino acid content of selected wild and domesticated cultivars of finger millet. *Plant Foods Hum Nutr* 1993;43:97–104. <http://dx.doi.org/10.1007/BF01087914>.
- [3] Dida MM, Srinivasachary, Ramakrishna S, Bennetzen JL, Gale MD, Devos KM. The genetic map of finger millet, *Eleusine coracana*. *Theor Appl Genet* 2007;114:321–2. <http://dx.doi.org/10.1007/s00122-006-0435-7>.
- [4] Upadhyaya HD, Pundir RPS, Gowda CLL. Genetic resources diversity of finger millet – A global perspective. In: Mgonja M, Lenne JM, Manyasa E, Srinivasaprasad S, editors. *Finger millet Blast management in East Africa. Creating opportunities for improving production and utilization of finger millet* Proceeding of the First International Finger Millet Stakeholder Workshop. ICRISAT; 2007 (196 pp.).
- [5] Upadhyaya HD, Sarma NDRK, Ravishankar CR, Albrecht T, Narasimhulu Y, Singh SK, et al. Developing a mini-core collection in finger millet using multilocation data. *Crop Sci* 2010;50:1924–31. <http://dx.doi.org/10.2135/cropsci2009.11.0689>.
- [6] Upadhyaya HD, Gopal Reddy V, Sastry DVSSR. Regeneration guidelines: Finger millet. In: Dulloo ME, Thormann I, Jorge MA, Hanson J, editors. *Crop specific regeneration guidelines* [CD-ROM] Rome, Italy: CGIAR System-wide genetic resource programme; 2008.
- [7] Mace ES, Buhariwalla HK, Crouch JH. A high-throughput DNA extraction protocol for tropical molecular breeding programs. *Plant Mol Biol Rep* 2003;21:459–60. <http://dx.doi.org/10.1007/BF02772596>.
- [8] Schuelke M. An economic method for the fluorescent labelling of PCR fragments. *Nat Biotechnol* 2000;18:233–4. <http://dx.doi.org/10.1038/72708>.
- [9] Liu K, Muse SV. PowerMarker: An integrated analysis environment for genetic marker analysis. *Bioinformatics* 2005;21:2128–9. <http://dx.doi.org/10.1093/bioinformatics/bti282>.
- [10] Excoffier L, Laval G, Schneider S. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evol Bioinformatics Online* 2005;1:47–50.
- [11] Perrier X, Jacquemoud-Collet JP. DARwin software. [cited October 10, 2014]. Available from Internet at: <http://darwin.cirad.fr/2006;2006>.
- [12] Billot C, Rivallan R, Sall MN, Fonceka D, Due M, Glaszmann JC, et al. A reference microsatellite kit to assess for genetic diversity of *Sorghum bicolor* (Poaceae). *Am J Bot* 2012;99:245–50. <http://dx.doi.org/10.3732/ajb.1100548>.
- [13] Generation Challenge Programme (GCP) Central Registry. [Cited October 15, 2014]. Available from Internet at: <http://gcpcr.grinfo.net/>.
- [14] International crops research institute for the semi-arid tropics (ICRISAT). [Cited October 15, 2014]. Available from Internet at: <http://www.icrisat.org/bt-molecular-geneflow.htm>.
- [15] Arya L, Verma M, Gupta VK, Seetharam A. Use of genomic and genic SSR markers for assessing genetic diversity and population structure in Indian and African finger millet (*Eleusine coracana* (L.) Gaertn.) germplasm. *Plant Syst Evol* 2013;299:1395–401. <http://dx.doi.org/10.1007/s00606-013-0822-x>.
- [16] Dida MM, Wanyera N, Harrison Dunn ML, Bennetzen JL, Devos KM. Population structure and diversity in finger millet (*Eleusine coracana*) germplasm. *Trop Plant Biol* 2008;131–41. <http://dx.doi.org/10.1007/s12042-008-9012-3>.
- [17] Manyasa EO, Tongona P, Shanahan P, Mgonja MA, De Villiers S. Genetic diversity in East African finger millet (*Eleusine coracana* (L.) Gaertn) landraces based on SSR markers and some qualitative traits. *Plant Genet Resour* 2014;1–11. <http://dx.doi.org/10.1017/S1479262114000628>.
- [18] Lule D, De Villiers S, Sewalem T, Dida M, Masresha, Kimani W, et al. Genetic diversity and eco-geographical distribution of *Eleusine* species collected from Ethiopia. *Afr Crop Sci J* 2014;22:45–58.