JARQ 44 (3), 259−268 (2010) http://www.jircas.affrc.go.jp

Genetic Relationships of Sorghum Germplasm in Asia and Africa Revealed by Rice cDNA-STS and Indel Markers

Pjotr STRELCHENKO[1], Hisato OKUIZUMI[2], Tariq SHEHZAD[3], Elena MALINOVSKAYA[1], Makoto KAWASE[2*], Shuichi FUKUOKA[2] and Kazutoshi OKUNO[3]

[1] All-Russian Research Institute of Plant Industry (42 Bolshaya Morskaya Street, 190000 St. Petersburg, Russia)
[2] Genebank, National Institute of Agrobiological Sciences (Tsukuba, Ibaraki 305−8602, Japan)
[3] Graduate School of Life and Environmental Sciences, University of Tsukuba (Tsukuba, Ibaraki 305−8572, Japan)

Abstract

The elucidation of *Sorghum bicolor* intraspecific genetic diversity is of interest in studies on evolutionary forces under domestication and geographical distribution, and has application in the design of programs for conservation, management and uses of genetic resources in breeding programs. Since molecular markers enable assessment of genetic relationship, various types of markers have been developed in many investigations of crops. This work aimed to assess the level of genetic diversity in grain sorghum collected from different regions in Asia including Central Asia and Africa. For the purpose, 14 rice cDNA-STS and 10 insertion/deletion (Indel) markers have been used. The majority of genes tagged by these markers were genes of complex crucial enzymes of cellular metabolism and stress-inducible proteins. Based on the markers, 48 sorghum accessions selected from the collection in NIAS Genebank (Tsukuba, Japan), and those from the All-Russian Research Institute of Plant Industry (St. Petersburg, Russia) were analyzed using cluster and principal coordinate analyses. The results demonstrated the existence of three major geographical centers for sorghum evolution and differentiation. Two of the centers are found in the African continent where sorghum is suggested to have been domesticated. Asian sorghums are grouped in a geographical cluster distinctive from the African groups. More precisely, there are separate sub-groups of accessions originating from East Asia and Central Asia, respectively. This is the first report of the Central Asian sorghum cultivars forming a unique genetic group. The variety groups demonstrated here may provide important information on phylogenic differentiation of this crop in Asia and Africa. This work demonstrated that rice cDNA-STS and Indel markers are useful instruments for evaluation of sorghum genetic diversity.

Discipline: Genetic resources
Additional key words: genetic diversity, grain sorghum

Introduction

Sorghum bicolor (L.) Moench is the most important domesticated species in the genus. It is grown as a staple food throughout Asian and African regions and as a forage and fodder crop for livestock in the remaining countries[6]. Understanding crop diversity is of interest in studies of evolutionary forces under domestication and geographical distribution, and will be applicable in the design of programs for the conservation, management and uses of genetic resources in breeding programs[13].

Classification of the most important agronomic form *S. bicolor* ssp. *bicolor* was established based on morphological characteristics of panicle and spikelet, which is rather complicated and included five major and 10 intermediate races[8].

Since molecular markers are an excellent tool for the assessment of genetic relationships, various types of markers have been developed in many studies in sorghum[1, 11, 17]. However, the majority of the studies have indicated that racial classification of sorghum is not always associated with genetic differentiation and genetic

* Corresponding author: Kawase@affrc.go.jp
Received 27 October 2008; accepted 10 December 2009.

diversity revealed by molecular studies. For example, Dahlberg et al. (2002) demonstrated that clusters developed by agronomic descriptors could not correspond with the groups classified by RAPD markers. Ritter et al. (2007) also illustrated the difficulty in verifying the sorghum races based on AFLP markers.

In contrast, Deu et al. (2006) investigated RFLP diversity in a set of 205 cultivated sorghum accessions in relation to geographical origin and racial differentiation. They identified 10 clusters using neighbor-joining analysis of the data, in which two main geographical groups were recognized, one primarily including accessions from north equatorial Africa and another containing accessions from south equatorial Africa. Asian accessions were distributed between those two main geographical groups. The authors suggest that the presence of Asian accessions in these two major geographical groups suggests two introductions of sorghum into Asia in its crop evolution. The majority of the clusters included a single predominant race, but accessions of races guinea and caudatum correspond with four and two clusters, respectively. Interestingly, a single cluster was composed only of the Chinese sorghums.

Rice is one of the world's most important food plants and has important syntenic relationships with other cereal species[12, 16] and is a model plant for the grasses. Campbell et al. (2007) demonstrated extensive gene sequence similarity in evolutionarily distinct species within the *Poaceae* family such as *Oryza sativa* and *Sorghum bicolor*. Extensive efforts have been undertaken to obtain detailed genetic and physical maps of the rice genome and the entire genomic sequence of rice has been published[9]. This information now is available in the Rice Annotation Project Database (RAP-DB, http://rapdb.dna.affrc.go.jp/). Then, it is possible to search DNA markers effectively in the research field of rice and its related species. In this work we tried to apply rice cDNA-STS and Indel markers, aiming to assess the level of genetic diversity in grain sorghum, especially in subspecies *bicolor*, originated from different regions of Asia including Central Asia and Africa.

Materials and methods

Of the 48 sorghum landraces studied (Table 1), 36 accession (1-36 in Table 1) originated from different regions of Asia and Africa, were sourced from the collection in NIAS Genebank (Tsukuba, Japan), and the remaining 12 accessions originating from the Central Asian region were sourced from the All-Russian Research Institute of Plant Industry (St. Petersburg, Russia). From the entire collection of 48 accessions 7, 11 and 13 accessions originated from East South and Central Asia, respectively. Seventeen accessions from different African countries were also studied.

Plant leaf material from a single plant per accession was dried using silica gel and 10-15 mg of the dried material was grinded into 2-ml Eppendorf tubes with a zirconium ball on a mixer mill. Genomic DNA was extracted using a micro scale CTAB method[7]. Extraction buffer (750 µl) (1% CTAB, 50 mM Tris-HCl pH 8.0, 10 mM EDTA, 0.7 M NaCl, 0.1% SDS, 0.1 mg /ml Proteinase K, 2% insoluble PVP and 2% 2-mercaptoethanol) was added. The homogenate was vortexed and incubated at 65°C for 60 min. Chloroform extraction was performed to remove cellular debris and proteins by addition of 750 µl chloroform-isoamyl alcohol (24:1 v/v), with shaking for 30 min and followed by centrifugation for 10 min at 10,000 rpm. DNA was precipitated by addition of 500 µl 2-isopropanol to the supernatant. After centrifugation (at 10,000 rpm for 10 min) the precipitate was washed twice in 70% ethanol. The final precipitate was dissolved in 50 µl of 1/10 TE solution containing RNase A, incubated at 42°C overnight, and stored at 4°C. The DNA concentration was measured on the NanoDrop ND-1000 (Thermo Scientific) spectrophotometer and diluted to a working concentration of 5 ng/µl using deionized water.

A standard PCR method was used to amplify DNA fragments from sorghum genome using primer pairs developed from the rice genome. In this study, "cDNA-STS" markers employed primer-pairs designed on rice cDNA clone, and "Indel" markers employed primer-pairs designed to contain an Indel polymorphism based on a database[15] (http://shenghuan.shnu.edu.cn/genefunction/ricemarker.htm). Twenty-four primer pairs are listed in Table 2. PCR conditions were optimized for each primer pair by adapting the annealing temperature (Tm). The PCR reaction performed consisted of 2 µl of DNA, 1 µl 10× PCR buffer, 1 µl dNTPs, 0.2 units *Taq* polymerase (TAKARA BIO INC), 1.5 µl mixture of two primers (10 pM/µl each), and 4.5 µl water in a total reaction volume of 10 µl. PCR cycling conditions were: an initial denaturation at 98°C for 3 min; followed by 30 cycles of 10 s at 98°C, 30 s at either 60°C or 55°C and 30 s elongation at 72°C, followed by a final elongation of 7 min at 72°C. All PCR reactions were performed on an iCycler (Bio-Rad). PCR products were analyzed by electrophoresis on 3% agarose gels (Sigma) in 0.5× TBE buffer (44 mM Tris-borate pH 8.3, 1 mM EDTA) at 150 V for approx. 1.5 h. Amplified fragments were visualized and sized using ethidium bromide (0.5 µg/ml) staining. Analysis of the PCR products with a small size difference was carried out on the HDA-GT12 analyzer

Table 1. List of sorghum accessions studied

#	Region	Subregion	Acc. #*	Local/cultivar name	Country
1	Asia	East	91317	NUO GAO LIANG	China
2			48458	AI HUI	China
3			119461	TOKIBI	Japan
4			119448	TAKAKIMI	Japan
5			45423	HIMEKI ZAIRAI	Japan
6			45437	MOCTAC LOCAL	Korea
7			137757	72−8−13 (Pingdon, Taiwan) Waxy	Taiwan
8		South	48530	ALLAKH	Bangladesh
9			377	BATTANBAN	Cambodia
10			297	MARIANGARIJORA MUDDAHIHAL	India
11			401	RABI YANGAR JORA MITHUGADUR	India
12			48881	PI 229486 VULGARE	Iran
13			480	HAZERA 6014	Israel
14			48466	AS 5781 HUAN SA PHAUNG AH LPYSU	Myanmar
15			48532	JUNELO	Nepal
16			48531	EC 18868	Nepal
17			137786	87−9−21−3−2 (Pakistan)	Pakistan
18			119513	COL/PAK/1991/IBPGR/2724(2)	Pakistan
19		Central	49005	PI 220636 Q 2/3/56	Afghanistan
20			830	DZHUGARA	China (West)
21			5098	DZHUGARA	Pakistan (North)
22			638	DZHUGARA	Turkmenistan
23			1120	DZHUGARA RED	Turkmenistan
24			1278	DZHUGARA WHITE	Turkmenistan
25			1299	DZHUGARA NONDROOPING	Turkmenistan
26			1149	AK−DZHUGARA	Uzbekistan
27			1296	DZHUGARA WHITE (MATKHAIR)	Uzbekistan
28			1313	SORGHUM RED	Uzbekistan
29			2313	DZHUGARA WHITE	Uzbekistan
30			2361	DZHUGARA	Uzbekistan
31			4531	TEKE DZHUGARA KIZIL	Uzbekistan
32	Africa		48543	MN 401	Algeria
33			260	MORABA 74	Ethiopia
34			48612	GIZA 3/59	Ethiopia
35			490	AKLMOI WHITE	Kenya
36			81250	PI 152748 C	Kenya
37			76744	MAKHOTLONG I	Lesotho
38			76742	TENANT WHITE	Lesotho
39			48550	SCHROCK	Morocco
40			48544	AIT BRAHIM	Morocco
41			48545	CODY	Morocco
42			48546	KOURNIANIA	Morocco
43			48759	KA 24	Nigeria
44			522	EAR FROM PIETESBURG DL/60/107	South Africa
45			519	AW 70/12 DL/59/1532	South Africa
46			48631	E 37	Tanzania
47			48615	UGANDA L1	Uganda
48			48619	S. VULGARE 72−728−1	Uganda

*Acc.#: NIAS Genebank catalogue (stock#); underlined - Russian Genebank catalogue.

Table 2. Name of markers, its localization in the rice genome, marker type (cDNA-STS or Indel) and sequences of primer pairs selected for the analysis of 48 sorghum accessions

#	Marker name	Chr.	Marker type	Forward	Reverse
1	E20660	1	cDNA-STS	CCAAAGCCGAGGAGAAGAAG	AAGGCCTTCTGGTTCATGAG
2	R494	1	cDNA-STS	TGGAGACGTCGTCTGACCTC	AACGAGAGAACATCCCCTCG
3	S13528	1	cDNA-STS	AGCACATTCGGCAGTTTCTC	GCACGCCATACAACAGCCTC
4	S13654	1	cDNA-STS	CAGACACGCCAGGTTTGATG	AGAATGCCACACCAACAATG
5	18-49	1	Indel	GATGGTTTGGCTCCGTGGTA	CGCCAAAACTTTTCGTCTCG
6	C53493	2	cDNA-STS	AGTGGGGGAAGGAGACGGAC	TTGTGCACGGGCTTAATCAG
7	E3902	2	cDNA-STS	TTGCTGCTACTGCGAAGAAG	TGGATACAGAGATGCACGAG
8	C63279	3	cDNA-STS	TCAAGTCCATCCAGATCACC	TGAGACAACAGGTTTTACCG
9	E522	3	cDNA-STS	CTCGTGTCAAAATCGGCGTG	CAGCCTTGACGTCGTTGTTC
10	E11757	3	cDNA-STS	AGCATCCTCCTCTCCGACTC	AAAATAGCAAGGCAAAGTGG
11	06-42	3	Indel	TTCAAACCTCAGTGCTGCAA	CTCCGTAGGCCTTGGTGTTA
12	C11112	4	cDNA-STS	CCAGCAACAGGGGATGAAGC	CAGGCATAAAACGGAGTGGC
13	C52717	5	cDNA-STS	CCGTGGATTCAAGGTCTAAG	AACATTGTCTTCTTGCCCAC
14	10-33	7	Indel	TCGTAGCAGTGATCGTGAGG	TGATATGCTTTCCCGACACA
15	C61344	8	cDNA-STS	CGCTGATTATGAGAGTGGTG	AGATGCTTGAGGGCTTGAAC
16	15-19	8	Indel	TCGATCGATCCAGTCCCAAA	CGCGGCATCATTTATGAACA
17	07-14	9	Indel	CCATGATCAAACCACACAGC	TGTCAGGGCACCATGACTTA
18	07-19	9	Indel	TATACGCGAGCGCTCTTACC	TGTATTCGGATGTTGCCAAG
19	C30194	11	cDNA-STS	TGTATGAAGGACATATGCCC	AAACTACACACATCCAAACC
20	03-14	11	Indel	GTCTTCGCCGCCACCTTCCA	CGCCGCCCCCGAATCGAG
21	07-33	11	Indel	AAACTTCCCAACCCTCCAAC	GCGCAACAAACGAACAGATA
22	RZ869	12	cDNA-STS	TATCAATCACCCCCAACCTC	TTTTGGTATTTGCTGCATGG
23	12-17	12	Indel	ACCGTAGCGTTAGCATGGAC	ACTACGAGAATGCGGTGCTT
24	12-13	12	Indel	CAGAGAAATGGAAGGCATTTGG	TGACGAGCGTAGTTGCATGTC

(eGene Inc) using the manufacturer's protocol. The presence or absence of each fragment was coded either as 1 or 0, respectively, and scored in a binary data matrix. MultiVariate Statistical Package (MVSP), version 3.12a was used to estimate the similarity coefficient (Simple Matching) for all possible pair wise comparisons between accessions and used to perform the cluster (Unweighted Pair Group Method using Arithmetic Averages) and principal coordinate analyses.

Results and discussion

In total 166 primer pairs (128 cDNA-STSs and 38 Indel) developed from the rice polymorphism database were used in this work. Initially these markers were screened to select primer pairs which would amplify different PCR products on six sorghum accessions originating from different regions of Asia and Africa. The majority of the primer pairs either did not amplify any products, showed complicated patterns for the products or amplified fragments with the same patterns for all accession. Finally 14 cDNA-STS and 10 Indel markers were selected (Table 2), which produced simple patterns with one or a few bands and demonstrated polymorphism in sorghum. These markers were distributed over 10 out of 12 chromosomes of the rice genome (with the exception of chromosomes 6 and 10). Loci for all the markers were identified from the Rice Annotation Project Database (RAP-DB, http://rapdb.dna.affrc.go.jp/) using the primer pairs sequences. For the 24 markers selected, 20 markers in the rice genome were tags for 27 genes of proteins from different categories including 3 genes identical to known rice proteins, 12 genes similar to known rice protein, 6 genes from InterPro domain-

Table 3. Marker names, sizes of PCR products in the rice genome, sizes of PCR products scored in the analysis of 48 sorghum genotypes and characterization of studied loci in the rice genome

#	Marker name	Size in rice	Size in sorghum	Category*	Genes of proteins in the rice genome
1	E20660	224	250	II	Similar to 60S acidic ribosomal protein P2-B (CaRP2B).
2	R494	206	490, 500, 700	II	Similar to Cytochrome b5 reductase.
3	S13528	217	500	IV, III	III: RINGv domain containing protein.
4	S13654	171	350, 400, 1000	III	Appr>p cyclic nucleotide phosphodiesterase domain containing protein.
5	18-49	248	450	II	Similar to Cysteine synthase, mitochondrial precursor (EC 2.5.1.47) (O- acetylserine sulfhydrylase) (O-acetylserine (Thiol)-lyase) (CSase C) (CS-C) (OAS-TL C) (AtCS-C).
6	C53493	203	300	II	Similar to ATOZI1 protein (Stress-induced protein OZI1) (AT0 ZI1 protein).
7	E3902	213	700	II	Similar to Ethylene response factor 2 (Ethylene response factor 3).
8	C63279	188	200	III, I, I	III: Protein kinase-like domain containing protein. I: Low molecular mass heat shock protein Oshsp17.3. I: Low molecular mass heat shock protein Oshsp18.0.
9	E522	203	230, 260	II, II	II: Similar to Branched-chain-amino-acid aminotransferase 3, chloroplast precursor (EC 2.6.1.42) (Atbcat-3). II: Similar to Squalene monooxygenase 2 (EC 1.14.99.7).
10	E11757	176	490, 495	IV, I	I: Allene oxide synthase.
11	06-42	228	180	IV	
12	C11112	193	240, 250	II	Similar to Alternative oxidase 1a (Fragment).
13	C52717	202	220	II, II	II: Similar to 60S ribosomal protein L18. II: Similar to Ethylene receptor homolog.
14	10-33	236	140, 150	II	Similar to Thioredoxin reductase.
15	C61344	187	100	II	Similar to Tyrosyl-tRNA synthetase (Tyrosyl-tRNA ligase TyrRS). class-I aaRS.
16	15-19	228	350	III	DOMON related domain containing protein.
17	07-14	174	280	No	
18	07-19	150	500	iii	EGF-like calcium-binding domain containing protein.
19	C30194	211	250	NO	
20	03-14	175	400, minor	IV, IV	
21	07-33	192	350	III	SMAD/FHA domain containing protein.
22	RZ869	232	700, minor	V	
23	12-17	172	700	No	
24	12-13	159	450, minor	No	

* Category I: identical to known rice protein, II: similar to known rice protein, III: InterPro domain-containing protein, IV: conserved hypothetical protein, V: hypothetical protein.

containing proteins, 5 genes to conserved hypothetical proteins, and only 1 gene of a hypothetical protein (Table 3). These proteins are of importance for cellular metabolism, especially proteins involved in binding of ATP, FAD, NAD, zinc, and calcium. These are complex crucial enzymes that may be important for plant adaptation. Interestingly, there are some genes for stress-induced proteins such as heat shock and pathogenesis-related.

Two types of polymorphism were detected in the analysis. For 18 markers the presence or absence of the

same band was registered. Of the 24 markers, 6 (the sizes of DNA fragments for these markers are shown by underline in Table 3) demonstrated fragment size polymorphism. Primer pairs used amplified rice DNA products with sizes of 150-248 bp (Table 3). In a total of 32 bands of 100 bp -1000 bp were scored. Twenty six bands for 20 markers were scored in the single-major-band patterns that may suggest homology between definite loci in rice and sorghum, although often the fragments were longer in sorghum than in rice. The genome size of sorghum is about twice that of rice, which may suggest more frequent events of insertions and/or duplications in the evolution of the sorghum genome.

The obtained binary data matrix for 48 sorghum accessions based on the analysis of DNA polymorphism using rice cDNA-STSs and Indels is shown in Table 4. The frequency of bands varied from 0.10 to 0.98 with an average of 0.56. Each accession was characterized by 15 to 22 bands. Most of the accessions were distinguished with an exception # 19 and # 30 from Afghanistan and Uzbekistan, respectively. The final binary data matrix was used for construction of the similarity coefficient matrix. For all possible pair wise comparisons this coefficient varied from 0.39 to 1 with an average of 0.67. The most distant (coefficient 0.39) accession were # 48 from Uganda and # 41 from Morocco, # 48 from Uganda and # 32 from Algeria, # 46 from Tanzania and # 33 from Ethiopia, and # 11 from India and # 33 from Ethiopia.

The distribution of the 48 sorghum accessions on the dendrogram is represented in Fig. 1. All accessions combined into three clusters. Cluster 1 contained two accessions from Asia and four accessions from Uganda, Tanzania and Kenya. Cluster 2 was the largest and included 29 accessions: 26 originating from different regions of Asia and three African accessions (two from Morocco and one from Ethiopia). This cluster has a complex structure and is divided into five sub-clusters. However, there is a clear tendency for accessions originating from the same region of Asia to cluster together. Sub-clusters A-C contains 10 accessions including six accessions from South Asian countries. Sub-cluster D contained only accessions originating from the Central Asian region. Finally, the sub-cluster E combined mainly accessions from East Asia including Japan, China and Korea. Lastly, in cluster 3, there were 13 accessions: 10 from African and three from Asian countries.

The principal coordinate analysis demonstrated that the distribution of accessions is very similar to the above mentioned. On the plot of the two first principal axes (Fig. 2), three clear groups of accessions were identified. The first group contained five accessions which were distributed in cluster 1: # 47 and # 48 from Uganda, # 46 from Tanzania, # 9 from Cambodia and # 15 from Nepal. In the second group 27 accessions were combined and except for # 40 from Morocco, all accessions originated from Asia. Accessions from both East and Central Asia had a tendency to group together and this group corresponds to cluster 2 based on the composition of accessions. The third group combined 11 accessions from African countries and three accessions from Asia, including # 13 from Israel. On the plot two accessions from Kenya were situated between groups 1 and 3. Differences in the position of accessions from groups 2 and 3 were determined by axis 1, but groups 1 and 3 were determined by both axes. In total 35.2% of variability was detected by the first two axes (22.4% - by axis 1 and 12.8% - by axis 2).

Thus, based on the analysis of sorghum Indel polymorphism detected by rice cDNA-STS and Indel markers using two different methods of multivariate statistics (cluster and principal coordinate analyses) a very similar distribution of accessions was revealed. In this analysis, three major groups of genotypes were identified, which may correspond to the 3 major geographical poles for sorghum evolution and differentiation.

The African accessions formed the most distant two groups and were comprised of group 1 or the East African (Uganda and Tanzania) accessions and group 3 was comprised of the other African countries. Two accessions from Kenya were in an intermediate position between these groups. At the first order of differentiation between the accessions of groups 1 and 3 shows evidence for both the lowest means of similarity coefficient between accessions in pairs # 48 Uganda - # 41 Morocco, # 48 Uganda - # 32 Algeria, # 46 Tanzania - # 33 Ethiopia and the most distant position on the dendrogram and two dimensional plot of the principal component analysis. This observation suggests that there are two major diversity centers in African sorghums and is supported by the results reported by Deu et al. (2007) based on the analysis of the sorghum core collection with RFLP probes. It is probable that such bipolar evolution of sorghum occurred in different regions of Africa that could have contributed to isolation of gene pools and divergent evolution[4, 6].

Most of the Asian accessions used were clustered into a distinctive group 2, although a few Asian accessions were found in groups 1 and 3. Thus, only three African accessions were found in a total of 29 accessions in group 2. Clearer grouping of Asian accessions was demonstrated by principal coordinate analysis: group 2 on the plot contained 26 Asian and only one

Table 4. The final binary data matrix for 48 sorghum accessions based on the analysis of DNA polymorphism using rice cDNA-STSs and Indels (the first row: name of markers and size of band scored in bp)

Accession	E20660, 250	R494, 490	R494, 500	R494, 700	S13528, 500	S13654, 350	S13654, 400	S13654, 1000	18−49, 450	C53493, 400	E3902, 700	C63279, 200	E522, 230	E522, 260	E11757, 490	E11757, 495	06−42, 180	C11112, 240	C11112, 250	C52717, 220	10−33, 140	10−33, 150	C61344, 100	15−19, 350	07−14, 280	07−19, 500	C30194, 250	03−14, 400	07−33, 350	RZ869, 700	12−17, 700	13−12, 450
1	1	0	1	0	1	1	0	0	1	0	0	0	1	0	0	0	0	0	1	1	0	1	0	0	1	1	1	0	1	1	1	1
2	1	0	1	1	1	1	0	1	1	0	0	0	1	0	1	0	1	1	0	0	0	1	0	0	1	1	1	0	1	0	0	0
3	1	1	0	1	1	1	0	0	1	0	0	0	0	1	1	0	1	1	0	1	1	0	0	0	1	1	0	0	1	0	1	1
4	0	0	1	1	1	1	0	1	1	0	1	0	1	0	1	0	1	1	0	0	0	1	1	0	1	1	1	0	1	1	0	0
5	0	0	1	1	1	1	0	1	1	0	0	0	1	0	0	1	1	1	0	0	0	1	0	0	1	1	1	0	1	0	1	0
6	1	1	0	1	1	0	1	1	0	0	0	0	1	0	0	1	1	1	0	0	1	0	0	0	1	1	1	0	1	0	1	0
7	1	1	0	1	1	1	0	1	1	0	0	0	1	0	0	1	1	1	0	1	0	1	1	0	1	1	0	1	1	0	0	0
8	1	0	1	0	1	1	0	0	1	1	1	0	1	0	0	1	1	0	1	0	0	1	1	0	1	1	1	0	0	0	1	0
9	1	1	0	1	1	0	1	1	0	1	0	0	1	0	0	1	1	1	0	0	1	0	1	1	1	1	1	0	1	0	1	1
10	1	1	0	1	1	1	0	1	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	1	1	1	1	0	0	0
11	1	0	1	1	1	1	0	0	1	0	0	0	0	1	1	0	1	1	0	0	1	0	1	0	1	0	1	0	1	0	1	0
12	1	1	0	1	0	0	1	1	0	1	1	0	1	0	0	1	1	1	0	0	1	0	0	1	1	1	1	0	1	0	1	1
13	1	1	0	1	1	1	0	1	1	0	1	1	1	0	0	1	1	1	0	0	0	1	1	0	1	1	1	1	1	0	1	1
14	1	0	1	1	1	1	0	0	1	0	0	0	1	0	0	1	1	1	0	0	0	1	0	0	1	1	1	0	1	0	0	0
15	1	1	0	0	1	1	0	0	1	0	1	1	0	1	0	1	1	0	1	0	0	1	0	0	1	1	0	1	1	1	1	1
16	1	0	1	0	1	1	0	1	1	0	1	1	1	0	0	1	1	0	1	1	1	0	0	0	1	1	0	1	1	0	1	1
17	1	1	0	1	1	0	1	0	0	0	0	0	1	0	0	1	1	1	0	0	1	0	1	1	1	1	1	0	1	0	1	1
18	1	1	0	0	1	1	0	1	1	0	1	1	0	1	0	1	0	0	1	1	1	0	0	0	1	1	0	1	1	0	1	0
19	1	1	0	1	1	1	0	1	1	0	1	1	0	1	1	0	1	0	1	1	0	1	1	0	1	1	0	1	1	0	0	0
20	1	1	0	0	1	0	1	1	1	1	1	1	1	0	0	1	1	1	0	1	1	0	1	0	1	1	1	0	1	0	1	1
21	1	1	0	0	1	1	0	1	1	0	1	1	1	0	0	1	1	0	1	1	0	1	1	0	1	1	1	0	1	1	1	1
22	1	0	1	1	0	0	1	0	0	0	1	0	1	0	0	1	1	1	0	0	1	0	1	1	1	1	0	0	1	0	1	1
23	1	1	0	1	1	0	1	0	0	0	1	1	1	0	0	1	1	0	1	0	0	1	0	1	1	1	0	0	1	1	1	1
24	1	1	0	0	1	1	0	0	1	0	1	1	0	1	0	1	0	0	1	1	1	0	0	0	1	1	0	1	1	0	1	0
25	1	1	0	1	1	1	0	0	1	0	1	1	1	0	0	1	1	0	1	0	0	1	0	0	1	1	0	1	1	0	0	0
26	1	1	0	1	1	1	0	0	1	0	1	0	1	0	1	0	1	1	0	0	0	1	1	0	1	1	1	0	1	0	0	0
27	1	1	0	0	1	1	0	0	1	0	1	1	0	1	0	1	1	0	1	1	1	0	1	0	1	1	1	1	1	0	0	1
28	1	1	0	1	1	1	0	1	1	0	1	1	1	0	0	1	0	1	0	0	0	1	1	0	1	1	0	0	1	0	0	0
29	1	0	1	1	1	1	0	1	1	0	1	0	0	1	1	0	1	1	0	0	0	1	1	0	0	1	0	0	1	0	1	1
30	1	0	1	0	1	1	0	0	1	0	1	1	1	0	0	1	1	0	1	1	1	0	1	0	1	1	1	1	1	0	0	1
31	1	1	0	0	1	1	0	0	1	0	1	1	0	1	0	1	1	0	1	1	1	0	0	0	1	1	0	1	1	0	1	0
32	1	1	0	1	1	1	0	0	1	0	0	0	1	0	1	0	1	1	0	0	1	0	1	0	1	1	1	0	1	0	0	0
33	1	1	0	0	1	1	0	0	1	0	1	0	1	0	0	1	1	0	1	1	0	1	1	0	1	1	1	0	1	0	0	0
34	1	1	0	0	1	1	0	0	1	0	1	1	1	0	0	1	0	0	1	1	0	1	0	0	1	1	0	1	1	0	1	1
35	1	1	0	0	1	1	0	0	1	0	1	1	0	1	0	1	0	0	1	1	0	1	1	0	1	1	1	1	1	0	0	1
36	1	0	1	1	1	1	0	1	1	0	1	0	1	0	1	0	1	1	0	0	0	1	0	0	1	1	1	0	1	0	0	0
37	1	0	1	1	1	1	0	1	1	0	1	0	1	0	0	1	1	0	1	0	1	0	1	0	1	1	0	0	1	0	0	1
38	1	0	0	1	1	1	0	0	1	0	1	0	1	0	0	1	1	0	1	0	1	0	1	0	1	1	0	0	1	0	0	1
39	1	0	1	1	1	1	0	0	1	0	1	0	1	0	0	1	1	0	1	0	1	0	1	1	1	1	0	0	1	0	1	0
40	1	1	0	1	1	1	0	0	1	0	1	0	1	0	0	1	1	1	0	0	1	0	1	0	1	1	0	0	1	0	0	1
41	1	0	1	1	1	1	0	1	1	0	0	0	1	0	0	1	1	0	1	0	1	0	1	0	1	1	0	0	1	0	0	0
42	1	1	0	1	1	1	0	0	1	0	0	0	1	0	1	0	1	1	0	0	1	0	1	0	1	1	1	0	1	0	0	0
43	1	0	1	1	1	1	0	0	1	0	1	0	1	0	0	1	1	0	1	0	1	0	1	0	1	1	1	0	1	0	1	1
44	1	0	1	1	1	1	0	1	0	0	1	0	0	1	0	1	1	0	1	0	1	0	1	0	1	1	1	0	1	1	0	0
45	1	1	0	1	1	1	0	1	1	0	1	0	1	0	0	1	1	1	0	0	1	0	1	0	1	1	1	0	1	0	0	1
46	1	0	1	1	1	1	0	0	1	0	1	0	1	0	1	0	1	1	0	0	1	0	1	0	1	1	1	0	1	0	1	1
47	1	0	1	1	1	1	0	0	0	0	1	0	0	1	0	1	1	0	1	0	1	0	1	0	1	1	1	0	1	1	0	0
48	1	0	1	1	1	0	1	0	1	1	1	0	0	1	0	1	1	1	0	0	0	1	1	0	1	0	0	0	1	0	1	0

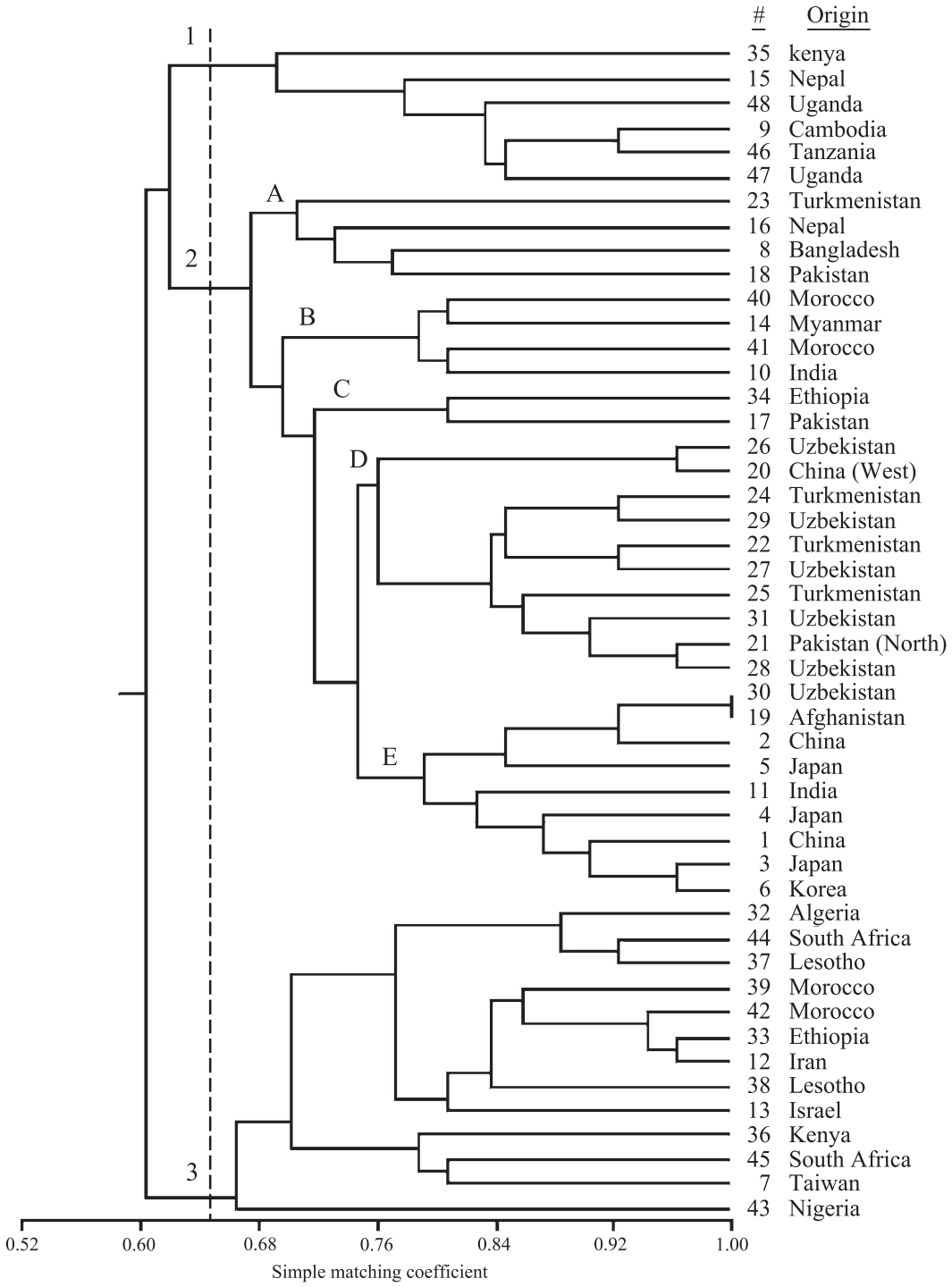

Fig. 1. Genetic relationships of 48 sorghum accessions

The dendrogram was constructed by MVSP software using simple matching coefficient on the bands amplified using rice cDNA-STS and Indel primer-sets and UPGMA algorithm on presence/absence score of each band.

African accession.

Our data contain some contradictions to the results of Deu et al. (2006). They identified only two major geographical poles in sorghum germplasm connected with northern and southern equatorial Africa. Some of the Chinese sorghum landraces are grouped in a cluster among 10 different clusters revealed based on RFLP analysis of 210 sorghum landraces. The majority of

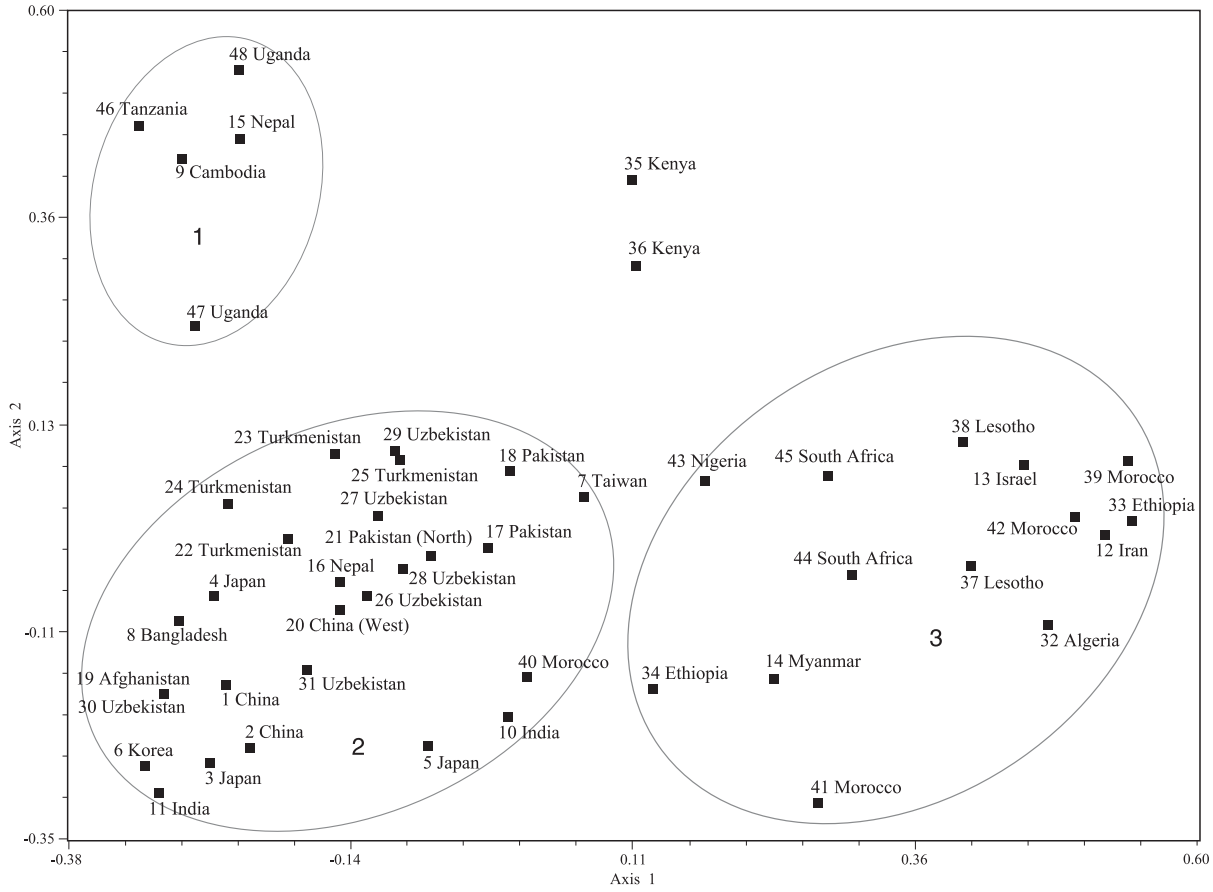

Fig. 2. The distribution of 48 sorghum accessions in principal coordinate analysis
Accessions were plotted on the first two principal coordinates (Axes 1 and 2).

Asian accessions were found together with African landraces in other clusters. Kimber (2000) suggests the domestication of sorghum occurred in eastern Africa 3,000-6,000 years ago with subsequent spread to the entire African continent and reaching Asia during only the first millennium (AD). The present results demonstrated availability in sorghum germplasm of the separate Asian geographical pole, which probably has a long period of independent evolution.

In this study, we aimed to assess the level of genetic diversity in grain sorghum collected from different regions in Asia including Central Asia and Africa. Based on the represented results, a complex structure of an Asian geographical diversity group was demonstrated. Three sub-groups were recognized in the Asian cluster: South Asian (clusters A-C), Central Asian (D) and East Asian (E). Accessions from the last two sub-groups had a clear tendency to group together on the plot of principal coordinate analysis. Deu et al. (2006) showed the genetic uniqueness of Chinese sorghums using RFLP markers. In our work, the East Asian sorghum cluster contains Chinese, Japanese and Korean accessions.

It was the first study of Central Asian sorghum using molecular markers. It was clearly demonstrated that these Central Asian accessions were characterized as unique genotypes, which formed a separate sub-group in the Asian group of sorghum. These materials will be used for further genetic studies and breeding works.

Conclusions

The results of our work demonstrated that recent achievements in rice genomics may be practically useful for the investigation of other crops in the *Poaceae* family. For instance, cDNA-STS and Indel markers might be successfully used for assessing the relationships between sorghum accessions and defining its intraspecies genetic structure. Three geographical poles, which evidently reflect the genetic differentiation of sorghum during its evolution and geographical distribution around the world, were identified. Two of these represent African sorghum. Our investigation concludes for the first

time, that Asian sorghums form a separate geographical pole which has a complex genetic structure. Thus, sorghum from both East and Central Asia form distinctive genetic groups amongst the Asian sorghums. This is the first report of sorghum accessions from Central Asia studied by molecular markers. A large number of accessions from Central Asia are conserved in the Russian Genebank, which will serve as unique genetic resources for further studies.

Acknowledgments

The senior author acknowledges the receipt of the Japan Society for Promotion of Science for Visiting Research Fellowship which enables this work to be undertaken, the staff of the NIAS Genebank for collegiality, technical assistance, materials, and help whilst undertaking this work. The authors thank Helen Hill (Centre for Plant Conservation Genetics, Southern Cross University, Lismore, Australia) for constructive review of the manuscript.

References

1. Bhattramakki, D. et al. (2000) An integrated SSR and RFLP linkage map of *Sorghum bicolor* (L.) Moench. *Genome*, **43**, 988−1002.
2. Campbell, M. A. et al. (2007) Identification and characterization of lineage-specific genes within the *Poaceae. Plant Physiol*., **145**, 1311−1322.
3. Dahlberg, J. et al. (2002) Comparative assessment of variation among Sorghum germplasm accessions using seed morphology and RAPD measurements. *Crop Sci*., **42**, 291-296.
4. de Wet, J. M. J. & Huckabay, J. P. (1967) The origin of *Sorghum bicolor* II distribution and domestication. *Evolution*, **21**, 787−802.
5. Deu, M., Rattunde, F. & Chantereau, J. (2006) A global view of genetic diversity in cultivated sorghums using a core collection. *Genome*, **49**, 168-180.
6. Doggett, H. (1988) *Sorghum*. Longman Scientific and Technical, London, pp.512.
7. Edwards, K., Johnstone, C. & Thompson, C. (1991) A simple and rapid for the preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Res*., **19**, 1349−1358.
8. Harlan, J. & de Wet, J. (1972) A simplified classification of cultivated sorghum. *Crop Sci*., **12**, 172-176.
9. International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. *Nature*, **436**, 793−800.
10. Kimber, C. T. (2000) Origins of domesticated sorghum and its early diffusion to India and China. *In* Sorghum: Origin, history, technology, and production, eds. Smith, C. & Frederiksen, R., John Wiley and Sons, New York, 3−98.
11. Kong, L., Dong, J. & Hart, G. E. (2000) Characteristics, linkage-map positions, and allelic differentiation of Sorghum bicolor (L.) Moench DNA simple−sequence repeats (SSRs). *Theor. Appl. Genet*., **101**, 438−448.
12. Moore, G. et al. (1995) Cereal genome evolution: Grasses, line up and form a circle. *Curr. Biol*., **5**, 737−739.
13. Okuno, K. et al. (2005) Plant genetic resources in Japan: Platforms and destinations to conserve and utilize plant genetic diversity. *JARQ*, **39**, 231−237.
14. Ritter, K. B. et al. (2007) An assessment of the genetic relationship between sweet and grain sorghums, within *Sorghum bicolor* ssp. *bicolor* (L.) Moench, using AFLP markers. *Euphytica*, **157**, 161-176.
15. Shen, Y.-J. et al. (2004) Development of genome-wide DNA polymorphism database for map-based cloning of rice genes. *Plant Physiol*., **135**, 1198−1205.
16. TIGR (2007) Syntenic blocks between rice and sorghum. *In* Rice Genome Annotation. http://www.tigr.org/tdb/synteny/sorghum/figureview_desc.shtml.
17. Yonemaru J. et al. (2009) Development of genome-wide simple sequence repeat markers using whole-genome shotgun sequences of sorghum (*Sorghum bicolor* (L.) Moench). *DNA Res*., **16**, 187−193.